

1
2
3
4
5
6
7
8 Bottom-Up Processes Dominate Early Word Recognition in Toddlers
9

10 Janette Chow^{1*}, Armando Q. Angulo-Chavira^{2*}, Marlene Spangenberg¹, Leonie Hentrup¹ and
11 Kim Plunkett¹

12 ¹Department of Experimental Psychology, University of Oxford, Anna Watts Building, Radcliffe
13 Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, UK

14 ²Laboratorio de Psicolingüística, Facultad de Psicología, Universidad Nacional Autónoma de México
15
16

17 *Joint First Authors

18 Corresponding Author: Prof Kim Plunkett, Department of Experimental Psychology, University of
19 Oxford, Anna Watts Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG,
20 UK, Tel: +44 (0) 1865 271398 kim.plunkett@psy.ox.ac.uk.

21 Orcid: <https://orcid.org/0000-0003-0216-7480>
22
23

Abstract

This study set out to investigate whether the ‘phonological onset preference effect’ often reported in adult studies using the visual world task (i.e., increased attention to an object that is phonologically-related to a spoken-target word, such as *boat-bear*) is also contingent upon toddler participants having sufficient preview time to inspect the picture stimuli. Picture preview is thought to support the activation of phonological codes which can then be matched to the phonological representations extracted from incoming speech signals and the picture stimuli, supporting the ‘phonological mapping hypothesis’. We found that both toddlers and adults were able to show an early phonological onset preference in short preview conditions, though, adults’ early phonological onset preferences in the short preview condition was extinguished by the presence of a semantic competitor, replicating previous adult findings (Huettig & McQueen, 2007). Removal of a semantic competitor reinstated the phonological onset preference effect under short preview conditions for adults. Our findings indicate that toddlers are driven more by bottom-up, phonological information when selecting a referent in a visual world task, as compared to adults who are more inclined to exploit top-down, semantic information when directing their attention to a visual object, especially when there is insufficient preview time. We propose that, when implicit naming is improbable in short-preview conditions, a phonological onset preference effect is driven by mapping on the visual-semantic levels, which is more susceptible to top-down influences.

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

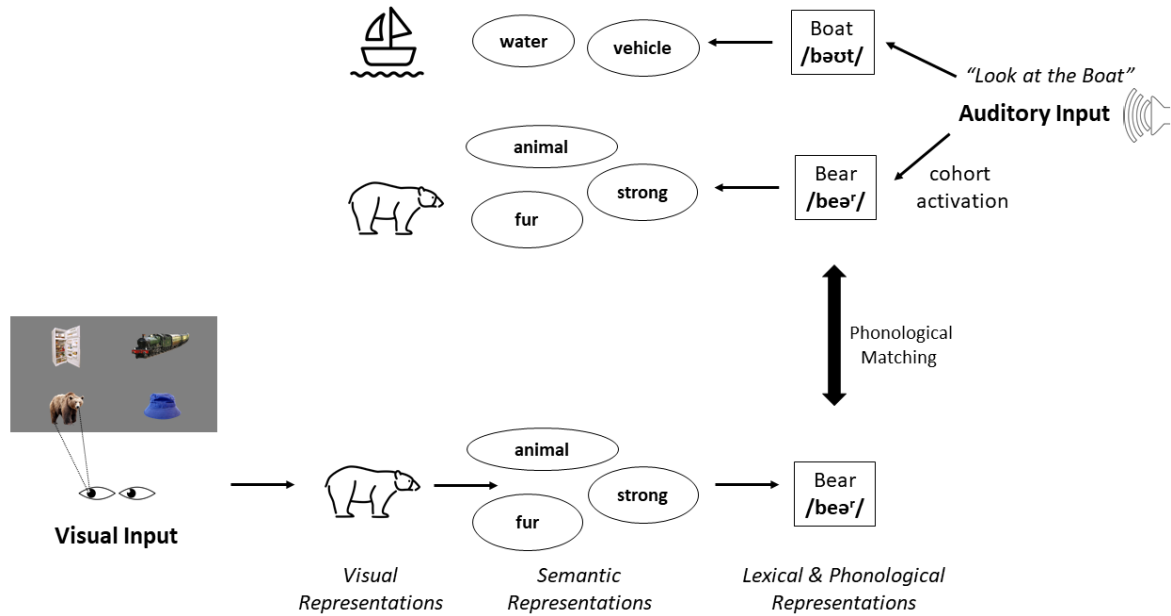
Bottom-Up Processes Dominate Early Word Recognition in Toddlers

Language processing involves integration of multisensory information. To comprehend the word “*duck*”, a toddler has to form an association between the speech sounds /d/ /ʌ/ /k/ and a duck’s perceptual features (e.g., has a bill, feathers, short legs, webbed feet, makes a quack noise, etc.), as well as semantic attributes such as a duck is a water bird like geese and swans. These associations are assumed to be stored in our mental lexicon, and become activated during speech processing to achieve word recognition and understanding. Upon hearing ‘Look at the duck in the pond’, if the referents are present (as in the visual world paradigm), the listener will process and integrate at least two streams of sensory information in parallel: extract meaning from the incremental speech input and direct visual attention to the relevant items in the visual scene.

This kind of language-mediated visual attention has been widely studied in the psycholinguistic literature using the Visual World Paradigm (VWP). In a typical VWP study, participants are allowed to look freely at a visual display (e.g., an array of four pictures), and hear a sentence containing a target word. Eye-tracking measures have demonstrated that our visual attention can be driven by a match (or partial match) between the auditory and visual input. For example, upon hearing the target word “*beaker*”, adult and toddler participants direct their eye gazes to an object that shares the same onset even in the absence of the target object (e.g., *beaver*) (Allopenna et al., 1998; Chow et al., 2017; Huettig & McQueen, 2007). Items that are semantically-related (even if not phonologically-related) to the spoken target word also attract attention (e.g., “*piano*” and *trumpet*) (Chow et al., 2017; Duñabeitia, Avilés, Afonso, Scheepers, & Carreiras, 2009; Huettig & Altmann, 2005; Johnson, McQueen, & Huettig, 2011; Yee, Huffstetler, & Thompson-Schill, 2011; Yee, Overton, & Thompson-Schill, 2009; Yee & Sedivy, 2006). Adults and toddlers are also sensitive to the perceptual attributes of items. Adults show increased attention towards objects that have a similar shape to the referent of the spoken-target word (e.g., “*snake*” and *rope*) (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2007; Huettig & McQueen, 2007). Both adults and toddlers will show increased attention towards objects that share the

same colour (e.g., “*spinach*” and *frog*) as the referent of the spoken target word (Huettig & Altmann, 2011; Johnson et al., 2011; Mani et al., 2012).

a) Phonological Matching Hypothesis



b) Visual-Semantic Matching Hypothesis

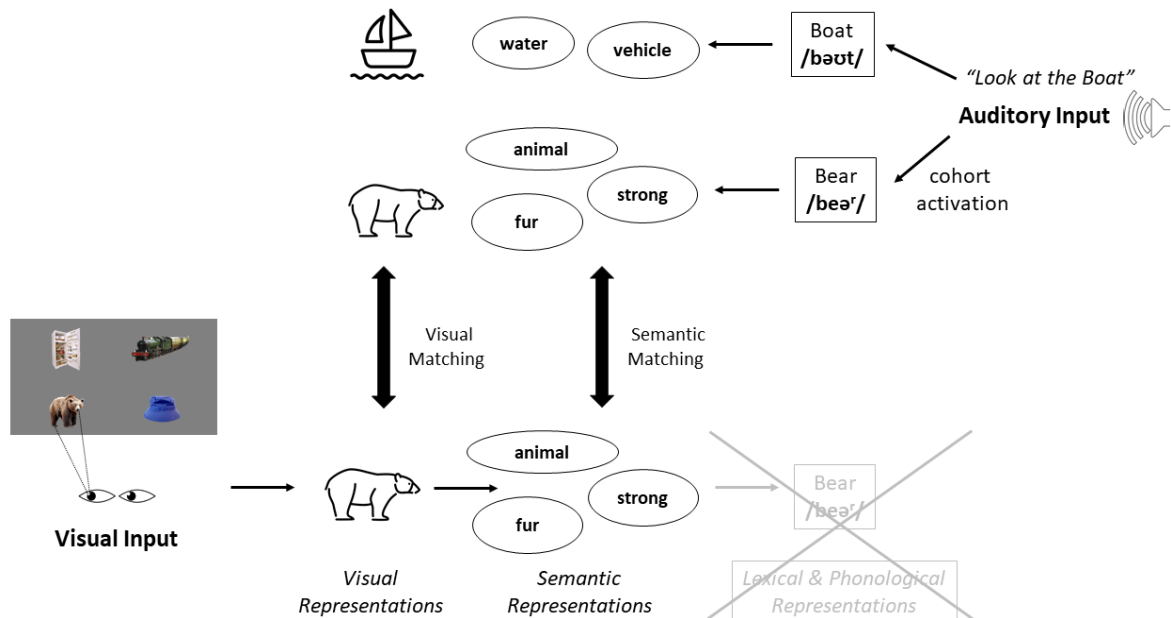


Figure 1¹. The visual-input route involves cascaded activation of visual, semantic and phonological representations. The auditory-input route involves cascaded activation of phonological, semantic and visual representations. Eye gaze is driven by the matching of representations. a) The phonological matching hypothesis assumes that viewing a picture (e.g., *bear*) would lead to the automatic activation

¹ Picture credits: Bear by Yu luck from the Noun Project. Boat by Graphic Tigers from the Noun Project. Eyes by Fahmihorizon from the Noun Project.

of not only its visual and semantic representations, but also, its lexical and phonological representations. A phonological onset preference would be driven by phonological matching. b) The visual-semantic matching hypothesis assumes that picture viewing would lead to the automatic activation of its visual and semantic representation, and therefore, a phonological onset preference would be driven by visual and semantic matching.

In order to achieve language-mediated visual attention, the incoming linguistic representations must be linked to the spatial location occupied by the picture in the visual display. How is this achieved? Take adults' and toddlers' preference for a phonologically-related item as an example (see *Figure 1*). From the *visual input*, participants process the visual form of the pictures, access its perceptual and semantic representations, i.e., recognising the picture's identity (Potter, 1975), and may eventually retrieve the picture's related lexical and phonological representations. Activation of the shared phonological onset of the spoken target and phonological competitor is considered to underlie the participant's phonological onset preference. The idea that looking at a picture leads to the automatic activation of a picture's lexical and phonological representations is not without controversy. While some theories posit that lexical and phonological representations are only retrieved when one is required to name the picture (as opposed to passive viewing) (see visual input route of *Figure 1b*) (Levelt et al., 1999), others have shown that processing of a target object can be slowed down by the mere presence of a phonologically-related competitor (see visual input route of *Figure 1a*) (McQueen & Huettig, 2014; Meyer, Belke, Telling, & Humphreys, 2007; Meyer & Damian, 2007; Morsella & Miozzo, 2002; Navarrete & Costa, 2005; though see Strijkers, Holcomb, & Costa, 2011 for a review of opposite effects).

From the *auditory input* (e.g., "*Look at the Boat.*"), participants first extract the phonological and lexical representations from the incremental speech signal, leading to a cascaded activation of its semantic and visual representations (Cummings et al., 2006; Huettig & McQueen, 2007). It is assumed that eye gaze towards the phonological competitor is jointly driven by the matching or overlapping of activations from both visual and auditory inputs (Ferreira & Tanenhaus, 2007; McQueen & Huettig, 2012; Tanenhaus et al., 2000). The coordinated interplay account (CIA) (Knoeferle & Crocker, 2006, cf. Altmann & Mirković, 2009) assumes that linguistic and visual processing are independent but tightly-related through processes of co-indexation and incremental reconciliation of the linguistic and

visual inputs. For example, visual world eye-tracking studies have shown that incremental syntactic interpretation (as sentences unfold auditorily) can direct visual attention towards objects mentioned in a visual display, and vice versa, contexts provided by the visual display can influence incremental syntactic interpretation (Knoeferle et al., 2005; Sedivy et al., 1999).

Based on the assumptions that linguistic and visual inputs are reconciled through co-indexation, various hypotheses concerning how activated representations are matched (or co-indexed) in order to drive eye gaze have been proposed. These hypotheses are all concerned with identifying the locus of the matching process. The ‘phonological mapping hypothesis’ proposes that fixations towards the phonological competitor are driven by the matching of the phonological representations activated from the visual and auditory inputs (Allopenna et al., 1998). For example, hearing /b/ not only activates the phonological representations of the target word “*beaker*”, but also the phonological onset competitor *beaver*, which matches with the phonological representations extracted from the picture of a *beaver*. The ‘semantic mapping hypothesis’ proposes that fixations towards a semantic competitor are driven by the matching of cascaded activation of semantic representations from the visual and auditory inputs (e.g., Huettig & Altmann, 2005). For example, hearing the target word “*piano*” leads to the activation of its semantic features ‘musical instruments’, which matches with the semantic features extracted from the picture of a *trumpet*. The ‘visual mapping hypothesis’ proposes that fixations towards a visual competitor are driven by a match of the activated visual representations from the visual and auditory inputs. For example, hearing the target word “*snake*” leads to the activation of its visual forms ‘long’ and ‘curved’, which matches with the visual features of a picture of a *rope*.

Theoretically, it is possible that all three hypotheses are correct, that language-mediated visual attention can be driven by the matching of representations on multiple levels. However, evidence suggests that some competition effects may only be explained by one of the mapping hypotheses. For example, Dahan and Tanenhaus (2005) and Huettig and Altman (2004) argue that the visual competitor effects reported in adults cannot be explained by the phonological- or semantic-mapping hypothesis as the authors controlled for phonological and semantic overlap between the target word and visual competitor (e.g. *snake* and *rope* do not sound similar and have different functions and category

membership). In contrast, a phonological competitor effect could in theory be explained by the visual- and semantic-mapping mapping hypothesis (*Figure 1b*), as well as the phonological hypothesis (*Figure 1a*): the target word can activate its cohort competitors which in turn can activate their associated semantic and visual representations, supporting potential matches at the semantic and/or perceptual levels. As illustrated in *Figure 1b*, an even stronger version of the visual mapping hypothesis denies that viewing a picture leads to the automatic activation of its phonological representation (Dahan et al., 2007; Dahan & Tanenhaus, 2005).

Findings reported by Huettig and McQueen (2007) led them to argue that the phonological competition effect found in their study is contingent on the matching of phonological representations alone, as predicted by the phonological mapping hypothesis. In their study (Experiment 1), adult participants saw a visual display containing a phonological competitor (e.g., *beaver*), semantic competitor (e.g., *fork*), a visual competitor (e.g., *bobbin*) and an unrelated distractor (e.g., *umbrella*), and heard a sentence containing a target word (e.g., “*beaker*”). When adult participants were given ample time (average 6.8 words) to preview the pictures prior to the acoustic onset of the target word, they showed a preference for the phonological competitor *early* in the trial and a preference for the semantic and visual competitors *later* in the trial. Preference for the semantic and visual competitors overlapped in the time course, likely due to the fact that there is a large overlap between semantic and visual properties of objects. For example, the perceptual features of a beaker (e.g., cylindrical) are also amongst its defining semantic features. The early phonological onset preference relative to the later semantic and visual preferences suggests that matching of representations occurs at different levels, roughly in the temporal order in which these representations have been activated from the visual and auditory inputs. On the assumption that viewing the visual display leads to a cascade of activations from perceptual to semantic and phonological activations, the extended picture preview before target word onset should enable the partial activation of the names of the objects in the display and the possibility of identifying the phonological competitor. Given that the only picture in the visual display which matches the onset of the unfolding spoken word is the phonological competitor, initial fixations should be directed to the phonologically matching object. As more of the target word is processed, the

1 phonological competitor can be rejected in favour of the semantic or visual matches. Huettig and
2 McQueen (2007) point out that if fixations of the objects in the visual display after target word onset
3 were based solely on the cascade of activations arising from the processing of the spoken word, there
4 would be no reason to suppose that fixations to the phonological competitor should occur *before*
5 fixations towards either the semantic or visual competitors² (*ibid* pg. 468).

6 In contrast, in Experiment 2 of their study, Huettig and McQueen (2007) report that the early
7 preference for the phonological competitor was abolished when picture preview time was drastically
8 reduced to 200ms. Under these conditions, adult participants exhibited a preference for the semantic
9 and visually related competitors alone. They argued that the lack of any phonological onset preference
10 in the short preview condition resulted from there being insufficient time for the adult participants to
11 generate the phonological codes of the visual stimuli. In the time taken for the target word to fully
12 unfold, the phonological competitor had already been dismissed as a candidate referent of the target
13 word. Such contrastive findings support the phonological hypothesis: fixations towards the
14 phonological competitor in Experiment 1 are driven by the cascaded activation of phonological
15 representations from the visual-input route (i.e., implicit naming of the pictures). Huettig and McQueen
16 (2007) concluded that "eye movements during language- mediated visual search depend on establishing
17 matches between information extracted from the visual display and from the speech signal. These
18 matches can be made at phonological, visual-feature and semantic levels of processing. Attentional
19 shifts thus appear to be co-determined by the type of information in the display (i.e., pictures or words),
20 the timing of cascaded processing in the word- and picture-recognition systems, and by the temporal
21 unfolding of information in the speech signal." (pg., 479).

22 It is also important to note that a phonological onset preference effect in the visual world task
23 is not contingent upon an extended picture preview of the candidate objects in the visual array before
24 the target word is heard. Task demands can also influence the impact of phonological competitors. In a
25 recent study, Apfelbaum, Klein-Packard, and McMurray (2021) used a 4-picture visual world task to

² It is also important to keep in mind here that Huettig and McQueen (2007) used a target-absent design where none of the pictures of the objects available in the display were an exact match for the target word.

1 monitor adult participants fixations of phonological competitors across a series of conditions. In a no-
2 preview condition, the stimuli consisted of a target word and 4 pictures. One picture depicted the target
3 (i.e., a target-present design), another picture depicted a phonological onset competitor to the target and
4 there were 2 unrelated pictures. Importantly, there was no semantic competitor nor a visual competitor.
5 Apfelbaum et al. (2021) report that participants looked more at the phonological competitor than the
6 unrelated pictures, thereby demonstrating that picture preview is not a prerequisite for observing
7 phonological competitor effects in the visual world task. Given the absence of any semantic or visual
8 competitors, this finding is readily understood in terms of the unfolding cascade of activations arising
9 from the processing of the spoken word: the *only* alternative match available (apart from the target
10 picture itself) is the phonological competitor. In contrast, in the short preview condition (Experiment 2)
11 of Huettig and McQueen (2007), there are semantic and visual competitors available to attract fixations,
12 permitting the phonological competitor to be quickly dismissed as the target word unfolds. A critical
13 difference between the no-preview condition in Apfelbaum et al. (2021) and the short preview condition
14 in Huettig and McQueen (2007) is the absence/presence of semantic/visual competitors. We suppose
15 that the semantic/visual competitors in the target-absent design used by Huettig and McQueen (2007)
16 suppressed any manifestation of a phonological competitor effect, precisely for the reason that Huettig
17 and McQueen (2007) suggested, namely, “there was not enough time for the retrieval of picture names
18 before the evidence in the speech signal was able to rule out the phonological competitor as a viable
19 lexical hypothesis in the word-recognition system.” (p.470). We predict that removal of the
20 semantic/visual competitors in a task, otherwise similar to Huettig and McQueen’s Experiment 2,
21 should reinstate the phonological competitor effect. We test this prediction in Experiment 3 of the
22 current manuscript.

23 24 *The Current Study*

25 This study set out to further our understanding of the mental representations and processes
26 involved in language-mediated visual attention in toddlers, and its course of development. Specifically,

we pitted the phonological mapping hypothesis and ‘visual-semantic³ mapping hypothesis’ against each other by manipulating the preview time, in a manner similar to Huettig and McQueen (2007). If the phonological hypothesis is correct, we should observe an early phonological onset preference in a long-preview condition, but not in a short-preview condition with toddlers. If the visual-semantic hypothesis is correct, we should observe an early phonological onset preference in both the long and short-preview conditions.

Based on the phonological-mapping hypothesis, we predicted that a long-preview time is required for both toddler and adult participants to extract the phonological codes from the visual display, and subsequently show an early fixation preference for the phonological competitor immediately after auditory target word onset. In contrast, we predicted an absence of an early phonological effect in the short-preview time condition in both toddler and adult participants, due to insufficient time to extract the phonological codes from the visual display prior to auditory target word onset. In Experiments 1 and 2, we attempted to replicate Huettig & McQueen’s (2007) findings with a set of fully-yoked, counterbalanced, and toddler-friendly realistic images used in Chow, Aimola Davies, and Plunkett (2017, Experiment 1). We presented 30-month-old toddler and adult participants with a four-picture display (a phonological onset competitor, a semantic competitor, and two unrelated distractors), and a sentence containing a target word, with either a short- or long-picture preview time. In Experiment 3, we further examined adults’ performance in a short preview condition when no visual-semantic competitor was available, by presenting them with a four-picture display that included two types of phonological competitors (onset and rhyme), and two unrelated distractors.

³ Note that it is very difficult to empirically differentiate between the semantic hypothesis and visual hypothesis as there is a great deal of overlap in the time course of semantic and visual competitor fixations in a VWP study (De Groot et al., 2016; Huettig & McQueen, 2007), due to the overlapping nature of semantic and visual features of objects. This situation is exacerbated in experiments involving infants and toddlers because of their limited vocabularies. We therefore refer to the ‘visual-semantic mapping hypothesis’ not because we believe the visual and semantic hypotheses are inseparable, but in the context of experiments with infants and toddlers the distinction is operationally mute.

Experiment 1

Experiment 1 examined the presence (or absence) of an early phonological onset preference in a short vs. long preview condition in 30-month-old toddlers. Chow et al. (2017) reported that similar to adults (Huetting & McQueen, 2007), 24- and 30-month-old toddlers showed an early phonological onset preference and a later semantic preference when given a long picture-preview. However, toddlers' language mediated visual attention has not been examined under a short-preview time condition. We predicted a replication of the finding in Chow et al. (2017, Experiment 1) for the long-preview condition. As existing findings suggest that toddlers generate lexical/phonological codes when viewing a picture in silence (Duta et al., 2012; Mani & Plunkett, 2010, 2011; Styles et al., 2015), we predicted that toddlers' preference for the phonological competitor is driven by a phonological match, and therefore, will only show an early phonological onset preference in the long-preview condition, not the short-preview condition.

Method

Participants

Twenty-five toddlers were included in this experiment. An additional five toddlers were excluded due to fussiness during testing, and, as a result, insufficient data was collected. All toddlers were recruited from monolingual English households. The mean age was 29.96 months (range= 27.65 – 31.66; 8 female, 17 male). Twenty-four toddlers took part in both the short- and long-preview conditions. One toddler took part only in the long-preview condition and did not take part in the short-preview condition after a break. The study was approved by the University of Oxford Medical Sciences Ethics Committee.

The sample size was chosen to be similar to previous research with a similar experimental design in toddlers ($n = 24$; Chow et al., 2017). We performed a sensitivity power analysis using the package SIMR (Green & MacLeod, 2016) in R. This analysis revealed that the sample size produced a statistical power of 70.9% for the main interaction among time-terms, Competitor, and Preview (see Results). Note that the statistical power produced by our analysis is a little less than usually used (80%). The lack of differences between preview conditions probably produces this low power (see Results). If

we consider the two-way interaction, the statistical power increases (85%). This second power analysis better describes the main results in this experiment.

Design

This study employs a within-participants design. Each participant was presented with two conditions, the short-preview and long-preview conditions. The presentation order of the preview conditions was randomised between participants. Following Huettig and McQueen (2007) and Chow et al. (2017), we employed a target-absent design to promote the number of fixations to competitors. In each condition, the participants saw the same 37 trials, including 24 target-absent test-trials, and 13 target-present filler-trials. The filler trials served to maintain the participants' attention, as they may lose interest in looking if the target word never matches any of the pictures. Participants in each condition were randomly assigned to one of four lists (see Tables A2 to A5 in Appendix A), which served to counterbalance for trial order and picture location. Each list started with two filler trials, and there was a filler trial after every two test trials.

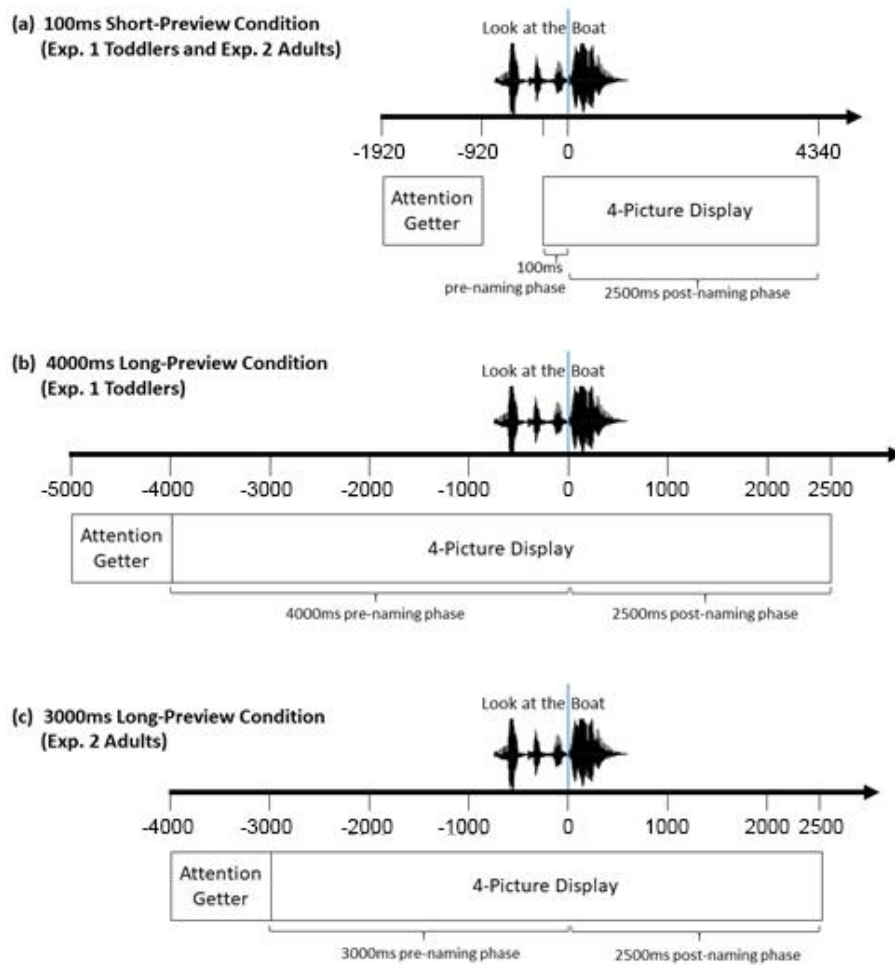


Figure 2. Timeline of a trial in the a) short-preview condition for adults and toddlers; b) long-preview condition for adults; and c) long-preview condition for toddlers.

In both, short- and long-preview conditions, the zero point was set to the onset of the target label. In the short-preview condition (*Figure 2a*), participants were presented with a 1000ms audio-visual attention getter at the beginning of the trial (from -1920 to -920ms relative to the target onset), followed by a four-picture visual display for a total of 2600ms (from -100 to 4340ms relative to the target onset). The target word onset, preceded by the prompt phrase “look at the” was always at 0ms. Therefore, the short-preview condition has a 100ms pre-naming phase and a 2500ms post-naming phase. In the long preview condition (*Figure 2b*), participants were presented with a 1000ms audio-visual attention getter at the beginning of the trial (from -4000 to -3000ms relative to the target onset), followed by a four-picture visual display for a total of 6500ms (from -100 to 4340ms relative to the target onset). The target word onset, preceded by the prompt phrase “look at the” was always at 0ms. Therefore, the

short-preview condition has a 100ms pre-naming phase and a 2500ms post-naming phase. The picture display appeared on-screen for a total of 6500ms (from -4000 to 2500ms relative to the target onset). The onset of the target label, preceded by the prompt phase “*look at the*”, was always at 0ms. Therefore, the long-preview condition has a 4000ms pre-naming phase, and a 2500ms post-naming phase.

Stimuli

The stimuli consisted of 24 picture of familiar objects used as visual competitors and 24 familiar words used as auditory target stimuli. From the 48 stimuli (24 pictures and 24 audios), 40 of the words included originated from the Oxford Communicative Development Inventory (OCDI) (Hamilton et al., 2000), the remaining eight were extra word norms collected alongside the OCDI since 2012. All words used were understood by 94.6% of the participants and produced by 89.3% (parental report). For each participant, all trials where either the target word or one of the picture labels was unknown were later removed for the analysis.

The 24 pictures were grouped into six sets of four pictures. Each yoked quadruple of pictures was presented four times (within-participant) so that, with a different target word (e.g. *boat*, *trousers*, *frog* and *hoover*⁴ respectively), each picture served once as the phonological competitor, semantic competitor, unrelated distractor 1 and unrelated distractor 2 (see Table A1 in Appendix A for more details). *Figure 3* shows a test trial with the target word “*boat*”, phonological competitor *bear*, semantic competitor *train*, and unrelated distractor 1 *fridge* and unrelated distractor 2 *hat*.

The four pictures in each set were visually dissimilar and unassociated according to the University of South Florida Free Association Norms (Nelson et al., 1998). Furthermore, the pictures were chosen so as to not resemble the referent of the target word regarding their colour or pattern (e.g., in a trial where the target word was “*tiger*” and the semantic competitor was a *butterfly*, the picture and colours were chosen so as to not have tiger-patterned- or black and orange coloured wings). The length of the 48 names of the competitor/distractor objects and the target words ranged from one to four

⁴ Hoover, instead of vacuum, is commonly used in British English.

1 syllables. All words belonged to one of eight categories (animals, clothing, electrical appliances, food,
2 furniture, toiletries, toys, vehicles).

3
4 **Visual Stimuli.** All pictures used were realistic photographs of the objects. Objects were edited out of
5 their original background and placed in the centre of a 9.53 x 9.53cm (50% grey) background using
6 Adobe Photoshop. The four square images (i.e., the areas of interest (AOIs) of eye-gaze measurement),
7 were located in the upper left, upper right, lower left and lower right corners of an invisible square, of
8 a total area of 28.58 x 28.58cm (23.7°), with a distance of 9.53cm (8.34°) separating the pictures
9 horizontally and vertically. All objects that had a typical colour were presented in this colour (e.g.,
10 ‘teddy’ in brown). All other objects were presented in a colour not present in their set so that all objects
11 in the same set had different colours. In each trial, we avoided presenting any of the objects in the
12 typical colour of or with a similar pattern as the target word. There was no obvious overall similarity
13 between the objects within the same set according to the researchers’ judgment. Note that the fully-
14 yoked design reduces any perceptual effect because it would be similar for each image in each picture
15 role. Pictures of animals and vehicles were always oriented so that their heads and fronts faced to the
16 middle in order not to cue looking into one particular direction.



17
18 *Figure 3.* An example display of a test trial with the target word "*boat*", a phonological competitor *bear*,
19 a semantic competitor *train*, and unrelated distractor 1 *fridge* and unrelated distractor 2 *hat*. Each
20 quadruple of pictures was yoked four times so that, with a different target word, each picture served
21 once as the phonological competitor, semantic competitor, unrelated distractor 1 and unrelated
22 distractor 2.

The picture location *per se* and the picture location by role was counterbalanced within participants as well as between participants: The location of each picture was different for all four blocks, so that every picture appeared in all of the four possible locations on the screen, exactly once for each condition for each participant and thus twice overall. All four locations were used equally often for a particular picture in a particular trial across all participants. Each picture role (Phonological Competitor, Semantic Competitor, Unrelated Distractor 1 and Unrelated Distractor 2) appeared equally often in each of the four possible locations for each participant. Thus, each role appeared 24 times on the right and 24 times on the left, and each role occupied each position nine times.

Auditory Stimuli. The auditory label (target word) presented with each picture set was preceded by the prompt phrase “*Look at the...*”. The sentences were recorded from a female British English speaker in child-directed speech. The audios were filtered and volume-matched using Goldwave. The mean duration of the prompt phrase was 558.41ms and the mean duration of the target word was 823.00ms. The audios were aligned so that the target word of each sentence would begin 100ms after picture onset in the short-preview and 4000ms in the long-preview conditions. In the short-preview condition, the sentence onset was at least 100ms after the offset of the attention getter, and the mean duration of the between stimulus interval (between attention getter and prompt phrase onset) was 361.59 ms.

Procedure

Toddler participants sat on their parent’s lap in front of a Tobii TX 300 eye tracker’s screen (23 inches, 1920 x 1080 resolution). The distance between the eyes of the participants and the screen was approximately 65 cm. The eye tracker was directly below the screen to track the participants’ eye movements. The study was run on Matlab using a custom stimuli presentation framework (PresentMate) 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. Before running the experiment, a nine-point calibration was performed with a colourful beach ball being the attention getter. Individual points were recalibrated in order to improve calibration when necessary. During the whole experiment, the background colour of the screen remained 50% grey. The auditory stimuli were presented through centrally-located

loudspeakers. A video camera above the screen allowed the experimenter to monitor the scene from the adjacent preview room. Each participant took part in the short- and long-preview conditions. The block order of the conditions was randomised between participants. Between the two blocks, there was a short break of around five minutes in which the toddler was given some time to play and have a drink. In both conditions, the experimenter initiated all trials individually via a computer key; a trial was only initiated when the participant was looking at the screen. Parents were instructed not to talk to the toddlers and to keep their eyes closed during the testing.

Data Processing

Pre-processing of eye-tracking data was conducted using a custom-made script in Matlab2016a. Minimum fixation duration was set to 100ms within a radius of 35 pixels. Aggregation of eye-tracking data and track loss analysis were conducted using the eyetrackingR package (Dink & Ferguson, 2015). To reduce the autocorrelation in the signal, fixation data between the auditory onset of the target word and the end of the trial was aggregated into twenty-five 100-ms timebins (i.e. 12 x 8.3 frames within each 100 ms timebin) for each picture type, condition, and participant. All statistical analyses were carried out using R 3.5.3 in RStudio.

Results

In the short-preview condition, data were available from 24 toddlers. In total, 505 out of 576 test trials were included in the analysis (88%). We excluded trials that contained words unknown to the participants (either the target word or one of the four picture labels), leading to the removal of 20 trials (3%). Fifty-one trials (9%) were excluded due to insufficient data (> 25% track loss) after target word onset. Thus, the average contribution of each toddler was 20 out of 24 test trials in the short-preview condition. In the long-preview condition, data were available from 25 toddlers. In total, 462 out of 600 test trials were included in the analysis (77%). Exclusions due to a lack of vocabulary knowledge amounted to 38 trials (6%). Nineteen trials recorded no data due to the participant not looking at the screen during the trial (3%), and a further 81 trials were excluded due to insufficient data (>25% track loss) after target word onset (14%). Thus, the average contribution of each toddler was 19 out of 24 test trials in the long-preview condition.

Preliminary examination of the overall patterns of fixations for the short- and long-preview condition (*Figure 4*) revealed that in both conditions, toddler participants showed an earlier preference for the phonological competitor over the unrelated competitor. Toddler participants also showed a later preference for the semantic competitor over the phonological and unrelated competitors lasting until the end of the trial.

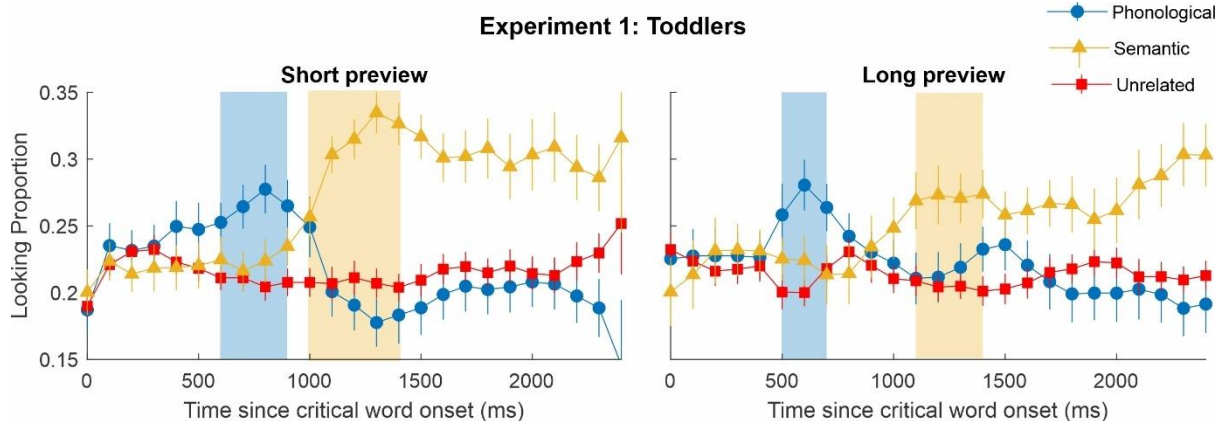


Figure 4. Fixation patterns in Experiment 1: toddler short-preview (100 ms) condition (left panel) and long-preview (4000 ms) condition (right panel). The points and lines represent experimental data with standard error bars. The shaded regions represent the significant clusters for phonological (blue) and semantic (yellow) competitors against unrelated ones.

Growth Curve Analysis

Previous works showed that phonological and semantic activation had different time courses (Chow et al., 2017; Huettig & McQueen, 2007). Recent adult VWP and infant IPL studies and the present research have used Growth curve analysis to describe the temporal dynamics of eye-tracking data (e.g., Barr, 2008; Law II & Edwards, 2014; Von Holzen & Mani, 2012). This analysis was carried out using mixed-effects binomial logistic regression models in R version 3.5.3 with glmmPQL, MASS 7.3-51.4. Binomial logistic models estimate the log of the odds ratio of the binary variable, which measures if participants fixated or not to the area of interest, coded by 1 and 0, respectively.

As it takes approximately 300ms to plan and launch an eye movement (Canfield et al., 1997), fixations within the first 300ms were unlikely to be driven by the target word. Therefore, we created two-time windows for analysis to better model the phonological and semantic effects. The first window measured phonological and semantic effects, and it analysed the period from 300 to 1400ms after presenting the auditory target onset. The second one explores only the early phonological effect, and it

analysed a shorter period from 300ms to 900ms relative to the target onset. These window sizes allow us to model the fixation with interpretable low-order polynomials (see chapter 3 in Mirman, 2014 for interpretation of orthogonal polynomials).

The dependent variable was the log odds ratio of fixations: $\log(\text{target fixation} / \text{non-target fixation})$. 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). Thus, the model has two levels. Level 1 estimates the change of fixation patterns over linear, quadratic, and cubic time trends. Level 2 estimates the effects of competitor (phonological onset vs. semantic vs. unrelated) and preview (short-preview vs. long-preview) on the change of fixation patterns on all time terms. Participants were included as a random effect on all time-terms. All fixed factors were dummy coded. The unrelated distractor (average of the two unrelated distractors) and the short-preview condition were treated as the baseline in the model; thus, relative parameters were estimated for the phonological onset and semantic competitors, and the long-preview condition. We used the following R formula to perform the analysis: $\text{fixations} \sim (\text{Linear} + \text{Quadratic} + \text{Cubic}) * \text{Competitor} * \text{Preview} + ((\text{Linear} + \text{Quadratic} + \text{Cubic}) | \text{Subject})$. Table 1 summarises the model statistics for toddlers (see *Figure 5* for the fitted model plots).

The results showed that toddlers' gaze had a faster linear increase in the semantic than the unrelated (Linear x Competitor: Unrelated-Semantic) and phonological (Linear x Competitor: Phonological-Semantic) competitors in the short preview. Furthermore, the phonological onset preference in the short-preview condition was modelled by the significant quadratic term (Quadratic x Competitor: Unrelated-Phonological; Quadratic x Competitor: Phonological-Semantic), indicating that toddlers had a more sharply increasing pattern of looks in the phonological than the semantic and the unrelated competitors. Thus, they looked faster at the phonological competitor and then looked away. Finally, toddlers' increase in looks to the semantic over the phonological competitor was faster in the short than the long preview condition according to the significant interaction with the linear time term (Linear x Preview: Short-Long x Competitor: Phonological-Semantic).

Table 1**Results for Experiment 1: Semantic and Phonological competition in toddlers**

Fixed factor	β	<i>SE</i>	<i>z</i>	<i>p</i>
(Intercept)	-1.310	0.021	-61.680	<0.001
Linear	-0.112	0.073	-1.539	0.922
Quadratic	0.099	0.073	1.352	0.974
Cubic	-0.062	0.073	-0.855	0.999
Competitor:Unrelated-Phonological	0.091	0.036	2.518	0.243
Competitor:Unrelated-Semantic	0.161	0.040	3.978	0.002
Preview:Preview:Short-Long	-0.029	0.030	-0.964	0.999
Linear x Competitor:Competitor:Unrelated-Phonological	-0.368	0.127	-2.883	0.094
Linear x Competitor:Unrelated-Semantic	0.822	0.121	6.775	<0.001
Linear x Competitor:Phonological-Semantic	1.191	0.142	8.352	<0.001
Quadratic x Competitor:Unrelated-Phonological	-0.488	0.127	-3.837	0.003
Quadratic x Competitor:Unrelated-Semantic	0.126	0.121	1.041	0.998
Quadratic x Competitor:Phonological-Semantic	0.615	0.142	4.328	<0.001
Cubic x Competitor:Unrelated-Phonological	0.148	0.126	1.173	0.994
Cubic x Competitor:Unrelated-Semantic	-0.045	0.121	-0.370	>0.999
Cubic x Competitor:Phonological-Semantic	-0.193	0.141	-1.369	0.971
Linear x Preview:Short-Long	0.042	0.106	0.393	>0.999
Quadratic x Preview:Short-Long	-0.160	0.107	-1.502	0.936
Cubic x Preview:Short-Long	0.044	0.106	0.414	>0.999
Preview:Short-Long x Competitor:Unrelated-Phonological	0.091	0.052	1.765	0.803
Preview:Short-Long x Competitor:Unrelated-Semantic	-0.031	0.051	-0.607	>0.999
Preview:Short-Long x Competitor:Phonological-Semantic	-0.122	0.058	-2.097	0.545
Linear x Preview:Short-Long x Competitor:Unrelated-Phonological	0.331	0.182	1.815	0.767
Linear x Preview:Short-Long x Competitor:Unrelated-Semantic	-0.531	0.176	-3.011	0.065
Linear x Preview:Short-Long x Competitor:Phonological-Semantic	-0.863	0.204	-4.228	<0.001
Quadratic x Preview:Short-Long x Competitor:Unrelated-Phonological	0.440	0.181	2.420	0.302
Quadratic x Preview:Short-Long x Competitor:Unrelated-Semantic	0.086	0.177	0.487	>0.999
Quadratic x Preview:Short-Long x Competitor:Phonological-Semantic	-0.354	0.203	-1.738	0.820
Cubic x Preview:Short-Long x Competitor:Unrelated-Phonological	0.064	0.181	0.354	>0.999
Cubic x Preview:Short-Long x Competitor:Unrelated-Semantic	-0.009	0.176	-0.056	>0.999
Cubic x Preview:Short-Long x Competitor:Phonological-Semantic	-0.074	0.203	-0.364	>0.999

Growth curve analysis using the formula for R function:

$\log(\text{Fixations/Non-fixations}) \sim (\text{Linear} + \text{Quadratic} + \text{Cubic}) * \text{Competitor} * \text{Preview} + ((\text{Linear} + \text{Quadratic} + \text{Cubic}) | \text{Subject})$.

Fixed factor: Competitor (Unrelated, Phonological, Semantic), Preview (Short: 100 ms; Long: 4,000 ms).

The analysis was performed over the period from 300 to 1400 ms after the presentation of the auditory label.

Bold rows indicate significant values.

1

2

The analysis of the phonological effect in the second window (Table 2) revealed that toddlers

3

looked faster to the phonological than the unrelated competitor in the short preview condition

4

(Competitor: Unrelated-Phonological; Linear x Competitor: Unrelated-Phonological).

Table 2**Results for Experiment 1: Phonological competition in toddlers**

Fixed factor	β	<i>SE</i>	<i>df</i>	<i>z</i>	<i>p</i>
(Intercept)	-1.086	0.041	836	-26.358	<0.001
Linear	0.165	0.094	836	1.762	0.078
Quadratic	0.035	0.094	836	0.374	0.708

Cubic	0.015	0.094	836	0.162	0.870
Competitor:Unrelated-Phonological	-0.200	0.048	836	-4.184	<0.001
Preview:Short-Long	-0.019	0.055	836	-0.354	0.723
Linear x Competitor:Unrelated-Phonological	-0.313	0.117	836	-2.663	0.007
Quadratic x Competitor:Unrelated-Phonological	-0.015	0.117	836	-0.127	0.898
Cubic x Competitor:Unrelated-Phonological	-0.016	0.117	836	-0.142	0.887
Linear x Preview:Short-Long	-0.063	0.137	836	-0.460	0.645
Quadratic x Preview:Short-Long	-0.202	0.137	836	-1.474	0.140
Cubic x Preview:Short-Long	-0.089	0.136	836	-0.656	0.511
Competitor:Unrelated-Phonological x Preview:Short-Long	-0.004	0.069	836	-0.062	0.950
Linear x Unrelated-Phonological x Preview:Short-Long	0.269	0.171	836	1.573	0.116
Quadratic x Competitor:Unrelated-Phonological x Preview:Short-Long	0.313	0.170	836	1.832	0.067
Cubic x Competitor:Unrelated-Phonological x Preview:Short-Long	0.119	0.170	836	0.700	0.483

Growth curve analysis using the R formula:
log(Fixations/Non-fixations)~(Linear+Quadratic+Cubic)*Competitor*Preview+((Linear+Quadratic+Cubic)|Subject).
Fixed factor: Competitor (Unrelated, Phonological), Preview (Short: 100 ms; Long: 4,000 ms).
The analysis was performed over the period from 300 to 800 ms after presenting the target label onset. Bold rows indicate significant values.

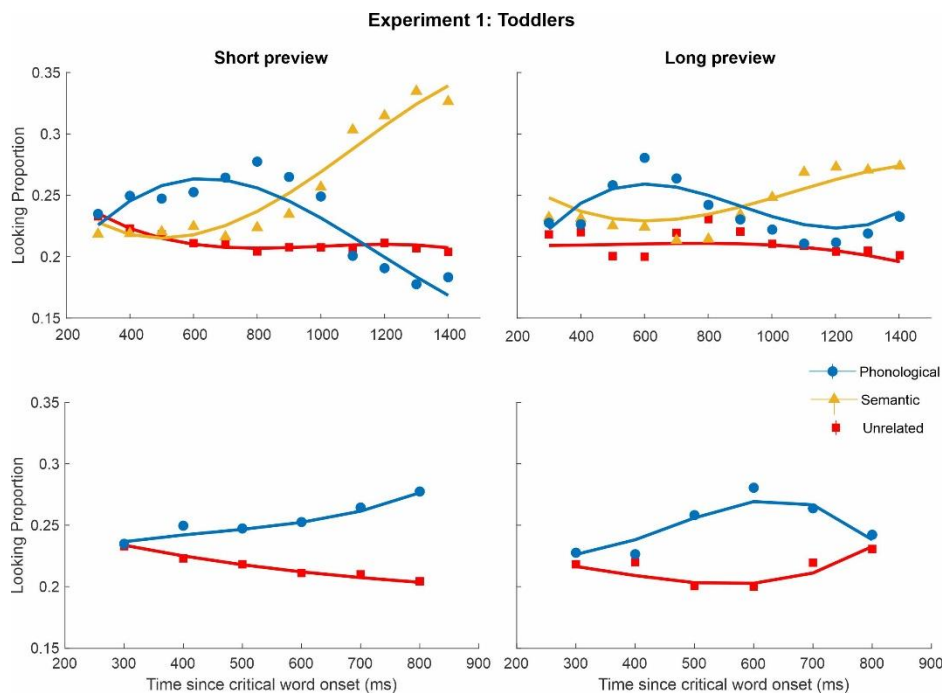


Figure 5. Modelled data for Experiment 1. Upper panels present the first growth curve analysis over the window from 300 to 1400 ms after presenting the target onset; this analysis explores semantic and phonological effects in the short (100 ms) and long (4000 ms) previews. Lower panels present the second growth curve analysis from 300 to 900 ms after presenting the target onset; this analysis only explores the phonological effects in short and long previews.

Cluster-based nonparametric analysis

We also applied the cluster-based non-parametric statistic (Maris & Oostenveld, 2007) to measure the temporal properties of the phonological and semantic effects: beginning, duration, and end.

This approach has been successfully used in studies of preferential looking times with toddlers (Angulo-Chavira & Arias-Trejo, 2021; Arias-Trejo et al., 2022). This method provides a non-supervised approach to detecting temporal segments in which time series are different between conditions using a permutation approach to create a null hypothesis. We analysed the binned (100ms) data from 300 to 1400ms (relative to the target onset) to explore phonological and semantic effects against unrelated competitors. To track differences between conditions, we used paired t -test. Clusters were created with the sum of adjacent t -values equal or greater than the critical values for the one-tail test. We used one-tail because we expected that participants to look more to the semantic and phonological competitors than the unrelated ones. A permuted distribution was created by randomly combining data between conditions (1,000 iterations). A cluster was considered significant using an alpha level of 0.05 regarding the permuted distribution.

Figure 4 shows the significant difference between the phonological and the semantic competitors against the unrelated distractors. They show that toddlers looked more to the phonological than the unrelated competitor; in the short preview from 600 to 900ms after the target onset ($t_{cluster} = 9.316$, $t_{max} = 2.849$, $p < 0.001$) and in the long preview from 500 to 800ms after the target onset ($t_{cluster} = 6.653$, $t_{max} = 3.035$, $p < 0.001$). Similarly, they looked more to the semantic than the unrelated competitor; in the short preview from 1000 to 1400ms after the target onset ($t_{cluster} = 21.273$, $t_{max} = 5.549$, $p < 0.001$) and in the long preview from 1100 to 1400ms after the target onset ($t_{cluster} = 9.963$, $t_{max} = 2.727$, $p < 0.001$).

In sum, both analyses showed that toddlers present an early phonological activation and a late semantic activation independently of the preview exposure time.

Discussion

In accordance with our expectations, Experiment 1 revealed that 30-month-old toddlers displayed a late semantic preference regardless of whether they were exposed to a short or long preview condition. However, contrary to our expectations, Experiment 1 also revealed that toddlers displayed an early phonological preference in both preview conditions. On the basis of findings with adult participants (Huettig & McQueen, 2007, Experiment 2), we predicted that toddlers would show an early

phonological onset preference in the long preview condition but *not* in the short preview condition. In fact, toddlers exhibited a very clear phonological onset preference in the short preview condition (see *Figure 5*).

These findings are inconsistent with the phonological mapping hypothesis, insofar as toddlers' early phonological onset preference is not contingent upon having a reasonably long preview time to extract the phonological codes of the visual display prior to auditory target word onset. While it is possible that toddlers are able to extract the phonological codes within 100ms of preview, previous research has reported that adults fail to show an early phonological onset preference with a short preview period (200ms) (Huettig & McQueen, 2007, Experiment 2). It seems counter-intuitive that toddlers are better able to extract phonological codes from the visual and auditory stimuli than adults. Furthermore, it is unlikely that the pattern of results reported in Experiment 1 is an artefact of the particular combination of stimuli used (Appendix A) since each object served as phonological competitor, semantic competitor and unrelated distractors across trials for every toddler in the study. A possible cause of the difference in behaviour of our toddler participants and Huettig and McQueen's adult participants is the character of the visual stimuli used in our experiment: We used toddler friendly, full colour, photographic images whilst Huettig and McQueen used line drawings to depict the objects in their experiment. In Experiment 2, we ask whether we can replicate the findings of Huettig & McQueen (2007) in adults using the same stimuli and a similar design to Experiment 1.

Experiment 2

Method

Participants

Twenty-four adults were included in this experiment. All adults were monolingual native British English speakers recruited from the University participant recruitment scheme. All adult participants reported normal or corrected-to-normal vision and no history of language or hearing impairments. Undergraduate students were awarded course credits, while non-students were

remunerated for their participation. The mean age was 21.31 years (range = 18.59 – 34.32; 19 female, 5 male).

The sample size was chosen to be similar to Experiment 1. The sensitivity power analysis (SIMR, Green & Macleod, 2016) revealed that the sample size produced a statistical power of 99.8% for the main interaction among time-terms, Competitor, and Preview (see Results), suggesting that Experiment 2 has sufficient statistical power.

Design

The general design and stimuli of Experiment 2 were identical to Experiment 1. As seen in *Figure 2c*, the long-preview condition has a 3000ms pre-naming phase (as opposed to 4000ms in Experiment 1⁵), and a 2500ms post-naming phase.

Procedure

The general procedures were identical to Experiment 1. Adult participants sat in front of the eye-tracking screen. They were instructed to listen to the sentences carefully, and that they can look freely on the screen but should not take their eyes off screen during the experiment.

Results

Data from all 24 adult participants were included in the analysis in short- and long-preview conditions. In the short-preview condition, a total of 573 out of 576 test trials were included in the analysis (99%). Three trials (1%) were removed due to insufficient data (> 25% track loss) after target word onset. Thus, the average contribution of each adult participant was 24 out of 24 test trials in the short-preview condition. In the long-preview condition, a total of 565 out of 576 test trials were included in the analysis (98%). Four trials recorded no data due to the participant not looking at the screen during the trial (1%), and a further seven trials were removed due to insufficient data (> 25% track loss) after target word onset (1%). Thus, the average contribution of each adult participant was 24 out of 24 test trials in the long-preview condition.

⁵ The preview time in Experiment 1 and Chow et al. (2017) was 4000ms for toddlers. To reduce boredom and waiting time in adults, we reduced the preview time to 3000ms in the long-preview condition.

The raw data for the short- and long-preview condition results are depicted in *Figure 6*. Adult participants showed a phonological onset preference only in the long-preview condition and a semantic preference in both the short- and long-preview conditions. Compared to toddlers, adults' phonological onset preference in the long-preview condition is much shorter, lasting no longer than 800ms, peaking at approximately 500 ms after the target onset.

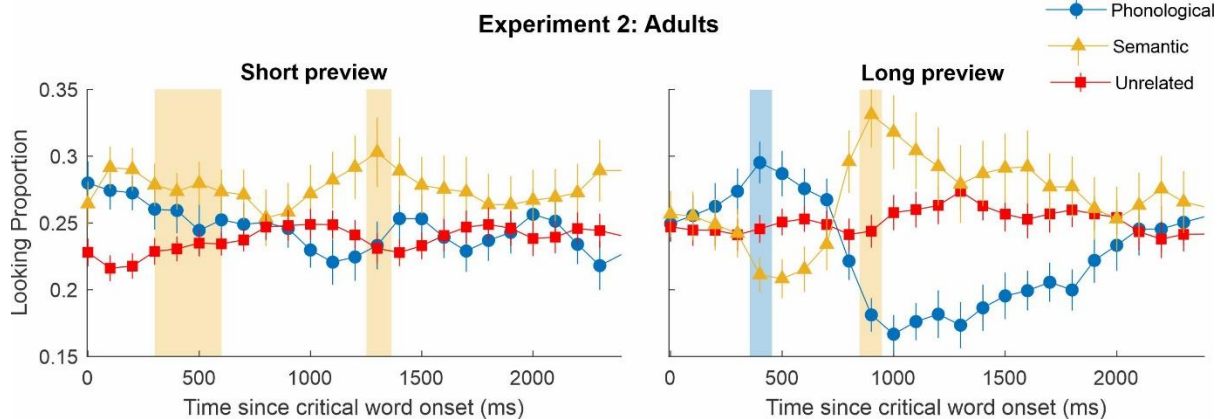


Figure 6. Fixation patterns in Experiment 2: short-preview (100 ms) condition (left panel) and long-preview (3000 ms) condition (right panel). The points and lines represent experimental data with standard error bars. The shaded regions represent the significant clusters for phonological (blue) and semantic (yellow) competitors against unrelated ones.

Growth Curve Analysis

The analysis strategy and parameters of the growth curve analysis were identical to Experiment

1. Table 3 summarises the model statistics (see *Figure 7* for the fitted model plots).

Table 3

Results for Experiment 2: Semantic and Phonological competition in adults

Fixed factor	β	SE	z	p
(Intercept)	-1.163	0.019	-59.178	<0.001
Linear	0.015	0.068	0.232	>0.999
Quadratic	-0.143	0.068	-2.102	0.541
Cubic	-0.051	0.068	-0.755	>0.999
Competitor:Unrelated-Phonological	0.031	0.033	0.937	0.999
Unrelated-Semantic	0.169	0.038	4.426	<0.001
Preview:Short-Long	0.077	0.027	2.798	0.119
Linear x Competitor:Unrelated-Phonological	-0.129	0.117	-1.103	0.997
Linear x Unrelated-Semantic	0.085	0.113	0.75	>0.999
Linear x Phonological-Semantic	0.214	0.131	1.634	0.88
Quadratic x Competitor:Unrelated-Phonological	0.257	0.117	2.2	0.461
Quadratic x Unrelated-Semantic	0.289	0.114	2.536	0.233
Quadratic x Phonological-Semantic	0.031	0.131	0.239	>0.999

Cubic x Competitor:Unrelated-Phonological	0.162	0.117	1.384	0.968
Cubic x Unrelated-Semantic	0.021	0.114	0.191	>0.999
Cubic x Phonological-Semantic	-0.14	0.131	-1.065	0.998
Linear x Preview:Short-Long	0.114	0.096	1.185	0.993
Quadratic x Preview:Short-Long	0.193	0.096	2.005	0.621
Cubic x Preview:Short-Long	0.07	0.096	0.728	>0.999
Preview:Short-Long x Competitor:Unrelated-Phonological	-0.231	0.049	-4.686	<0.001
Preview:Short-Long x Unrelated-Semantic	-0.106	0.047	-2.247	0.424
Preview:Short-Long x Phonological-Semantic	0.124	0.055	2.24	0.43
Linear x Preview:Short-Long x Competitor:Unrelated-Phonological	-0.92	0.169	-5.414	<0.001
Linear x Preview:Short-Long x Unrelated-Semantic	0.353	0.163	2.158	0.494
Linear x Preview:Short-Long x Phonological-Semantic	1.274	0.192	6.62	<0.001
Quadratic x Preview:Short-Long x Competitor:Unrelated-Phonological	-0.18	0.169	-1.061	0.998
Quadratic x Preview:Short-Long x Unrelated-Semantic	-0.571	0.163	-3.489	0.013
Quadratic x Preview:Short-Long x Phonological-Semantic	-0.391	0.192	-2.035	0.597
Cubic x Preview:Short-Long x Competitor:Unrelated-Phonological	0.2	0.169	1.183	0.993
Cubic x Preview:Short-Long x Unrelated-Semantic	-0.399	0.163	-2.437	0.291
Cubic x Preview:Short-Long x Phonological-Semantic	-0.599	0.192	-3.12	0.046

Growth curve analysis using the formula for R function:

$\log(\text{Fixations/Non-fixations}) \sim (\text{Linear} + \text{Quadratic} + \text{Cubic}) * \text{Competitor} * \text{Preview} + ((\text{Linear} + \text{Quadratic} + \text{Cubic}) | \text{Subject})$.

Fixed factor: Competitor (Unrelated, Phonological, Semantic), Preview (Short: 100 ms, Long: 3,000 ms). The analysis was performed over the period from 300 to 1,400 ms after presenting the target label onset. Bold rows indicate significant values.

1

2

3

4

5

6

7

8

9

10

11

12

13

The results of the growth curve analysis in the large time window (300 to 1,400 ms after the target label onset) showed that adults' gaze had a faster linear increase in the semantic than the unrelated competitors in the short preview condition (Competitor: Unrelated-Semantic). Furthermore, there was a greater difference between phonological and unrelated competitors in the long preview condition than in the short one (Preview: Short-Long x Competitor: Unrelated-Phonological). This difference was modelled by linear time terms indicating a faster decrease of looking behaviours in the long preview condition than the short one (Linear x Preview: Short-Long x Competitor: Unrelated-Phonological).

A negative quadratic term modelled the difference between semantic and unrelated competitors, indicating that the semantic competitor had a more sharply increasing pattern of looks than the unrelated one. Still, this pattern was more robust in the long than the short preview condition (Quadratic x Preview: Short-Long x Unrelated-Semantic). Finally, participants rapidly increased looks to the semantic competitor relative to the phonological one. This effect was greater in the long than the short preview

condition (Linear x Preview: Short-Long x Phonological-Semantic; Cubic x Preview: Short-Long x Phonological-Semantic).

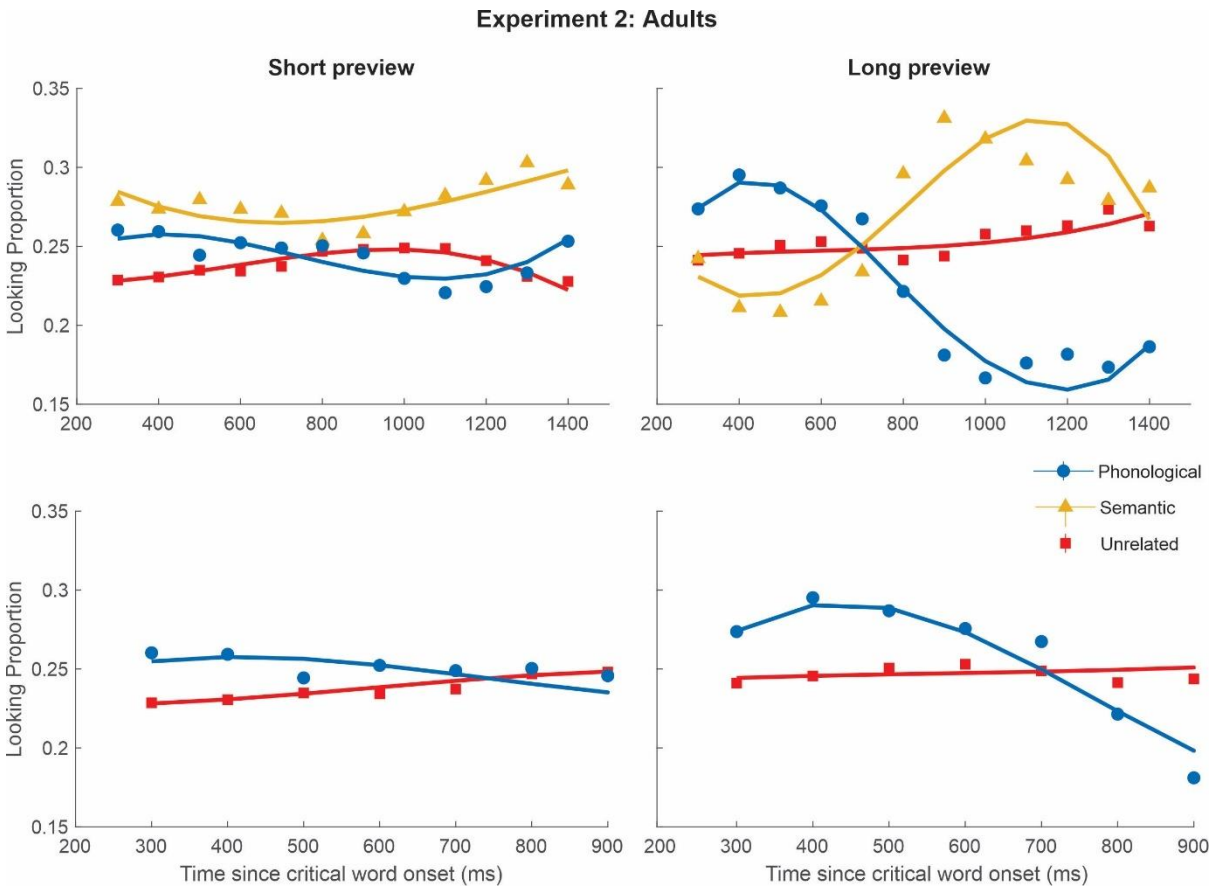


Figure 7. Modelled data for Experiment 2. Upper panels present the first growth curve analysis over the window from 300 to 1400 ms after the target onset; this analysis explores semantic and phonological effects in the short (100 ms) and long (3000 ms) previews. Lower panels present the second growth curve analysis over the 300 to 900 ms period after the target onset; this analysis explores only the phonological effects in short and long previews.

The second growth curve analysis for the early time window (300 to 900 ms) is summarised in Table 4 (see Figure 7). The results showed a greater difference between phonological and unrelated competitors in the long than the short preview condition (Preview: Short-Long x Competitor: Unrelated-Phonological). Participants looked faster to the phonological than the unrelated competitor. This effect was greater in the long preview condition than the unrelated one (Linear x Preview: Short-Long x Competitor: Unrelated-Phonological).

Table 4
Results for Experiment 2: Phonological competition in adults

Fixed factor	β	SE	df	z	p
(Intercept)	-1.13	0.03	1653	-35.1	<0.001
Linear	-0.11	0.1	1653	-1.15	0.249
Quadratic	0.110	0.09	1653	1.277	0.201
Cubic	0.109	0.09	1653	1.256	0.209
Competitor:Unrelated-Phonological	-0.03	0.03	1653	-1.03	0.304
Preview:Short-Long	-0.15	0.04	1653	-4.01	<0.001
Linear x Competitor:Unrelated-Phonological	0.128	0.11	1653	1.203	0.228
Quadratic x Competitor:Unrelated-Phonological	-0.26	0.11	1653	-2.42	0.015
Cubic x Competitor:Unrelated-Phonological	-0.16	0.11	1653	-1.52	0.128
Linear x Preview:Short-Long	-0.8	0.13	1653	-6.26	<0.001
Quadratic x Preview:Short-Long	0.011	0.13	1653	0.086	0.931
Cubic x Preview:Short-Long	0.271	0.13	1653	2.129	0.033
Preview:Short-Long x Competitor:Unrelated-Phonological	0.232	0.05	1653	5.152	<0.001
Linear x Preview:Short-Long x Competitor:Unrelated-Phonological	0.923	0.16	1653	5.951	<0.001
Quadratic x Preview:Short-Long x Competitor:Unrelated-Phonological	0.182	0.16	1653	1.174	0.240
Cubic x Preview:Short-Long x Competitor:Unrelated-Phonological	-0.20	0.16	1653	-1.29	0.196

Growth curve analysis using the formula for R function:

$\log(\text{Fixations/Non-fixations}) \sim (\text{Linear} + \text{Quadratic} + \text{Cubic}) * \text{Competitor} * \text{Preview} + ((\text{Linear} + \text{Quadratic} + \text{Cubic}) | \text{Subject})$.

Fixed factor: Competitor (Unrelated, Phonological), Preview (Short: 100 ms; Long: 4,000 ms). The analysis was performed over the period from 300 to 900 ms after presenting the target label onset.

Bold rows indicate significative values.

1

2 *Cluster-based nonparametric analysis*

3 The analysis strategy and parameters for the cluster analysis were identical to Experiment 1.

4 Statistically significant clusters are presented in *Figure 6*. There was no significant cluster between

5 phonological and unrelated competitors in the short preview condition. Adults looked more to the

6 semantic than the unrelated competitors in two-time windows, an early semantic effect from 300 to 600

7 ms after the target onset ($t_{cluster} = 7.803$, $t_{max} = 2.091$, $p < 0.001$), and a late semantic effect from 1300 to

8 1400 ms after the target onset ($t_{cluster} = 3.812$, $t_{max} = 1.976$, $p = 0.05$). Similarly, there were no differences

9 between phonological and unrelated competitors in the long-preview condition, but adults looked more

10 to the semantic than the unrelated competitor at 900 ms after the target onset ($t_{cluster} = 2.378$, $t_{max} = 2.378$,

11 $p = 0.02$). The cluster analysis did not replicate the phonological effect in the long preview condition

12 found in the growth curve analysis. In further exploration, we noted that the looking proportion in the

13 phonological competitor decreases dramatically after 900 ms in the long preview; therefore, the null

permuted distribution is biased by these values. To avoid this bias, we explore the phonological effect in a short time window from 300 to 900 ms after the target onset (comparable to the second growth curve analysis). This second analysis showed no significant differences between phonological and unrelated competitors in the short preview condition but a significant difference in the long-preview condition ($t_{cluster} = 1.976$, $t_{max} = 1.976$, $p = 0.02$).

In sum, both analyses revealed that the phonological effect was only observed when participants had a long preview; but the semantic effect was observed in both short and long previews.

Discussion

Experiment 2 demonstrated that adult participants displayed an early phonological onset preference in the long-preview condition, but not in the short-preview condition. Clearly, Huettig and McQueen's findings are unaffected by the replacement of line drawings with toddler friendly, full colour, photographic images. These findings replicate those of Huettig and McQueen (2007) who argue that, in the short-preview condition, there is insufficient time before target word onset for the adult participants to generate the phonological codes of the visual stimuli (implicit naming) and compare them to the target word. It is worth recapitulating here that, in principle, a phonological onset preference effect might be explained either by matching target word the cascading activations of visual, semantic and phonological representations arising from viewing the picture display (the phonological hypothesis), or by matching the pictures with the cascading activations of the phonological, semantic and visual representations arising from hearing the target word (the semantic-visual hypothesis). Huettig and McQueen (2007) argued that under the semantic-visual hypothesis, there would be no reason to suppose that fixations to the phonological competitor should occur *before* fixations towards either the semantic-visual competitors. However, in the long preview condition in this Experiment 2 (and their Experiment 1), adults demonstrated an early phonological onset preference, thereby supporting the phonological hypothesis. Furthermore, on the semantic-visual hypothesis, manipulation of the picture preview time should have little or no effect on the time course of competitor preferences. Clearly, this was not the case, neither in our Experiment 2, nor in Huettig and McQueen's Experiments 1 and 2. This pattern of

findings provides further support for the phonological hypothesis for precisely the reasons outlined earlier and by Huettig and McQueen (2007).

An important question remains unanswered: Why did adults fail to show an early phonological onset preference in the short preview condition whilst toddlers maintained this early preference under the same viewing conditions? As demonstrated in previous visual world studies, the matching process of language-mediated visual attention is highly dynamic and co-determined by various factors including the temporal unfolding of the auditory input, the visual stimuli that are present, and the timing of the cascaded processing of both the spoken-word and picture recognition processing systems (Huettig & McQueen, 2007; Huettig et al., 2012). In this study, participants had to extract and monitor two types of information (i.e., phonological and semantic) from the audio and visual stimuli. Specifically, in the short-preview condition where implicit naming is delayed, the temporal-relevance and strength of the phonological competitor diminish in comparison to the semantic competitor: the phonological competitor ceases to be relevant as soon as the target word has fully unfolded acoustically. Full activation of the target word (e.g., “*boat*”) would lead to lexical inhibition of its phonological competitors (including *bear*) (McClelland & Elman, 1986). Inhibitory effects within the mental lexicon are widely reported in adults and school-aged children, such as lexical competition and semantic inhibition effects (Dahan et al., 2001; Glaser & Dünghoff, 1984; Kapnoula et al., 2015; Schriefers et al., 1990; Weighall et al., 2016). In contrast, toddler studies have reported that phonological and semantic inhibitory mechanisms emerge and continue to develop towards the end of the second year (see General Discussion for further discussion) (Arias-Trejo & Plunkett, 2009, 2013; Chow et al., 2016, 2018; Mani & Plunkett, 2011). Moreover, for adults, the semantic competitor remains relevant to the spoken-target word as soon as the target word has been activated, and until the end of the trial. In other words, compared to toddlers, adults may be more inclined to exploit top-down, semantic information when directing their attention to a visual object. If temporal relevance and top-down, semantic information are the cause of the extinguished phonological onset competition effect in adults, we predict that removing the semantic competitor (or any strong competitor) should reinstate adults’ phonological onset preference effect in a short-preview condition. This prediction is tested in Experiment 3.

Experiment 3

In Experiment 3, we ask whether adults' lack of early phonological onset preferences in the short preview condition in Experiment 2 was due to their greater sensitivity to the temporal relevance and relative strengths of individual picture competitors, in particular the semantic-visual competitor, rather than a lack of preview time to generate phonological codes from the pictures alone. In this experiment, we replaced the semantic competitor with a phonological rhyme competitor (similar to Allopenna et al., 1998, but using a target-absent design), which typically leads to weaker and later effects than the phonological onset competitor (Allopenna et al., 1998; McQueen & Huettig, 2012). To allow direct comparison with Huettig & McQueen (2007, Experiment 2), the preview time was set to be 200ms.

Method

Participants

Thirty-one adults were included in this experiment. All adults were monolingual native British English speakers recruited from the University participant recruitment scheme. All adult participants reported normal or corrected-to-normal vision and no history of language or hearing impairments. Undergraduate students were awarded course credits, while non-students were remunerated for their participation. The mean age was 19.47 years (range = 18.52 – 22.62; 20 female, 11 male).

The sensitivity power analysis (SIMR, Green & MacLeod, 2016) revealed that the sample size produced a statistical power of 96.6% for the main interaction among time-terms, Competitor, and Preview (see Results), suggesting that Experiment 3 has sufficient statistical power.

Design

Each participant saw 41 trials, including 16 target-absent test-trials, and 21 target-present filler-trials. The target-present filler trials served to maintain the participants' attention, as they may lose interest in looking if the target word rarely matches any of the pictures. Participants in each condition were randomly assigned to one of four lists (see Table A3 in Appendix A), which served to counterbalance for trial order and picture location. Each list started with three target present trials.

Target absent trials never occurred back-to-back. Each rhyme competitor object also appeared as an onset competitor and as an unrelated distractor (apart from "foot" which never acted as a rhyme and "ship" which never occurred as an onset). Every time an object was repeated, a different image of this object was presented (i.e., there were three different images for each object). In order to minimise possible priming effects, there were always at least three trials (i.e., 14.76 seconds) in between trials that contained the same objects.

The target-absent test trials include two phonologically related objects: an onset competitor, which had a phonologically related onset to the auditory target word (i.e., *house* - *hand*) and a rhyme competitor, which rhymed with the target word (i.e., *house* - *mouse*). The two unrelated distractors were phonologically, semantically and visually unrelated to the target word. Target-present filler trials consisted of the target object (i.e., the referent) and three unrelated distractors which were phonologically, visually and semantically unrelated to the target object.

Procedure

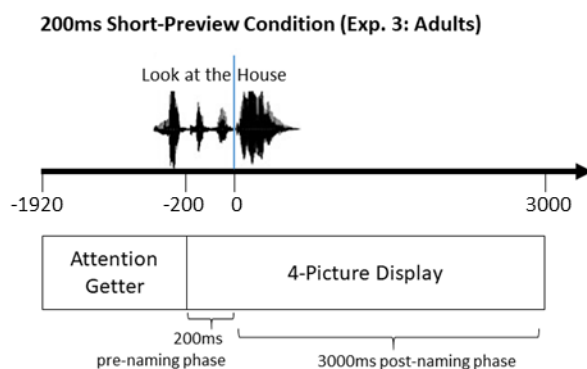


Figure 8. Timeline of a trial in Experiment 3.

Adult participants sat in front of the eye-tracking screen. They were instructed to listen to the sentences carefully, and that they can look freely at the screen but should not take their eyes off screen during the experiment. On each trial, participants were first presented with an attention getter in the middle of the screen. The onset of the target word was set up as the zero point as in previous experiments. The attention getter comprised of a rotating twirl that lasted for 1720 ms (from -1920 to -200ms relative

to the auditory target onset) and was accompanied by a sparkling sound lasting one second. A four-picture display appeared immediately after the offset of the attention getter and stayed on the screen for 3200 ms (from -200 to 3000 ms relative to the target onset). The visual display was accompanied by an auditory sentence instructing participants to look at the target object (e.g. "Look at the cat"). There was therefore a 200ms pre-naming phase, matching the short preview condition used by Huettig and McQueen (2007) and a 3000 ms post-naming phase (see *Figure 8*).

Results

Data from all 31 adult participants were included in the analysis. A total of 11 out of 496 test trials were excluded from the analysis (2%). The first four trials were removed for one participant due to technical failures (<1%), and a further seven trials (1%) were removed due to insufficient data (> 25% track loss) after target word onset. Thus, the average contribution of each adult participant was 15.6 out of 16 test trials.

The overall pattern of the results is depicted in *Figure 9*. Adult participants showed an early preference for the onset competitor lasting no longer than 1000ms after target word onset, peaking at approximately 700ms. Afterward, they showed an increased preference towards the rhyme competitor and the unrelated distractor.

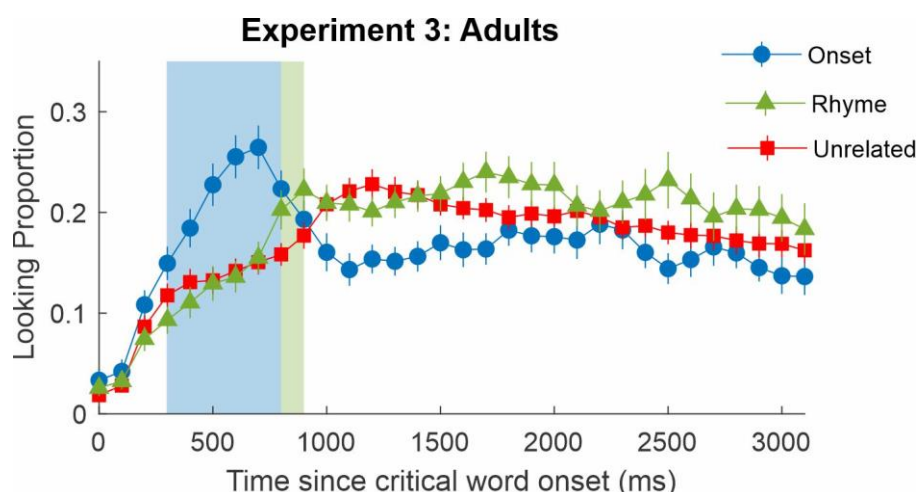


Figure 9. Fixation patterns in Experiment 3. The points and lines represent experimental data with standard error bars. The shaded regions represent the significant clusters for phonological onset (blue) and phonological rhyme (green) competitors against unrelated ones.

Growth Curve Analysis

Similar to the analysis performed in Experiment 1 and Experiment 2, we performed two analyses using different time windows. The first analysis modelled both onset and rhyme phonological effects from 300 to 1200 ms relative to the target onset. This window was selected because it captures the onset and rhyme phonological effects. This analysis included time terms (linear, quadratic, cubic) and competitor (onset, rhyme, unrelated) as fixed factors, and participants as a random effect on all time-terms (Formula: fixations~(Linear+Quadratic+Cubic)*Competitor+((Linear+Quadratic+Cubic)|Subject). Fixed factor: Competitor (Unrelated, Onset, Rhyme). Table 5 summarises the model statistics (see *Figure 10* for the fitted model plot).

The results of the first model showed that participants looked more to the onset-related than the rhyme-related (Competitor: Onset-Rhyme) and unrelated competitors (Competitor: Unrelated-Onset). The onset phonological effect was modelled by the quadratic (Quadratic x Competitor: Unrelated-Onset) and cubic (Cubic x Competitor: Unrelated-Onset) time terms. This effect reflects a faster and sharper increase of looks towards the phonological onset competitor compared to the unrelated one. This effect was similar when onset and rhyme competitors were compared (Quadratic x Competitor: Onset-Rhyme). Finally, the linear terms modelled the phonological rhyme effect, indicating that participants look faster to the rhyme competitor than the unrelated one (Linear x Competitor: Unrelated-Rhymes).

Table 5
Results for Experiment 3: Onset and Rhyme phonological competition in adults

Fixed factor	β	SE	z	p
(Intercept)	-1.83	0.092	-19.822	<0.001
Linear	0.387	0.106	3.623	0.004
Quadratic	-0.007	0.09	-0.078	>0.999
Cubic	0.062	0.089	0.696	0.999
Competitor:Unrelated-Onset	0.459	0.053	8.562	<0.001
Competitor:Unrelated-Rhymes	-0.048	0.059	-0.815	0.997
Competitor:Onset-Rhyme	-0.508	0.064	-7.894	<0.001
Linear x Competitor:Unrelated-Onset	-0.033	0.146	-0.226	>0.999
Linear x Competitor:Unrelated-Rhymes	0.652	0.158	4.106	<0.001
Linear x Competitor:Onset-Rhymes	0.685	0.174	3.937	0.001
Quadratic x Competitor:Unrelated-Onset	-0.525	0.144	-3.641	0.004
Quadratic x Competitor:Unrelated-Rhymes	0.119	0.158	0.752	0.998
Quadratic x Competitor:Onset-Rhyme	0.645	0.172	3.745	0.002
Cubic x Competitor:Unrelated-Onset	-0.062	0.142	-0.438	<0.001
Cubic x Competitor:Onset-Rhymes	-0.077	0.156	-0.498	0.999

Cubic x Competitor:Onset-Rhymes

-0.015

0.169

-0.092

>0.999

Growth curve analysis using the formula for R function:

$\log(\text{Fixations/Non-fixations}) \sim (\text{Linear} + \text{Quadratic} + \text{Cubic}) * \text{Competitor} + ((\text{Linear} + \text{Quadratic} + \text{Cubic}) | \text{Subject})$.

Fixed factor: Competitor (Unrelated, Onset, Rhyme).

The analysis was performed over the period from 300 to 1200 ms after the onset of the target label.

Bold rows indicate significative values.

The second model of Experiment 3 was performed over a window from 300 to 900ms (after the target onset) to capture only the phonological onset effect. The fixed factors were the three time terms (linear, quadratic, cubic) and the competitor (onset, unrelated). We only include the intercept of the participants as a random factor because the model failed to converge when the time terms were included (R formula: $\text{fixations} \sim (\text{Linear} + \text{Quadratic} + \text{Cubic}) * \text{Competitor} + (1) | \text{Subject}$)).

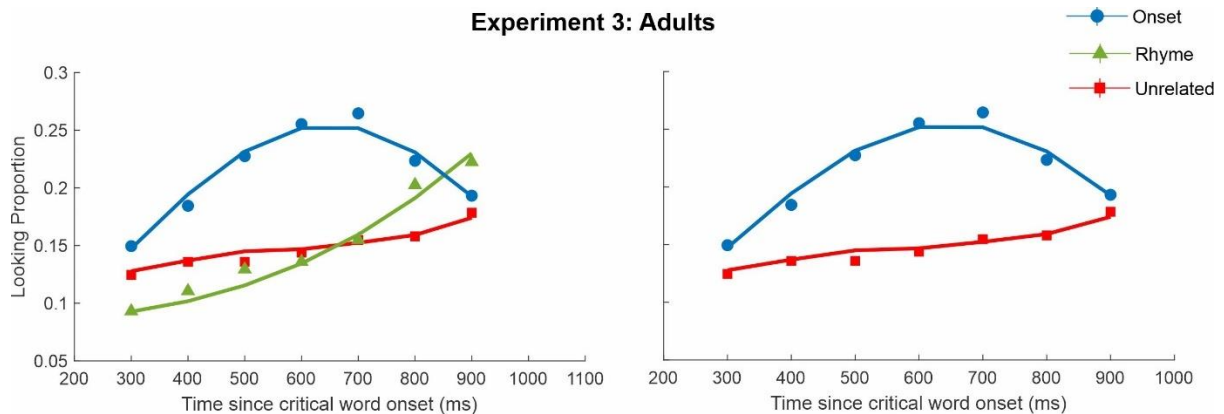


Figure 10. Modelled data for Experiment 3. The left panel presents the first growth curve analysis over the time window 300 to 1200 ms after the target onset; this analysis explores onset and rhyme phonological effects. The right panel presents the second growth curve analysis over the time window 300 to 900 ms after the target onset; this analysis explores only the phonological onset effect.

The results of the second model replicate those found in the first one. Participants had a greater looking proportion and a more sharply pattern of looks towards the onset phonological competitor than the unrelated one (Competitor: Unrelated-Onset; Quadratic x Competitor: Unrelated-Onset; Table 6).

Table 6

Results for Experiment 3: Onset phonological competition in adults

Fixed factor	β	SE	df	z	p
(Intercept)	-1.862	0.091	502	-20.249	<0.001
Linear	0.206	0.096	502	2.147	0.032
Quadratic	-0.042	0.095	502	-0.444	0.656
Cubic	0.015	0.095	502	0.164	0.869

Competitor:Unrelated-Onset	0.509	0.06	502	8.357	<0.001
Linear x Competitor:Unrelated-Onset	0.231	0.153	502	1.511	0.131
Quadratic x Competitor:Unrelated-Onset	-0.335	0.150	502	-2.222	0.026
Cubic x Competitor:Unrelated-Onset	-0.088	0.149	502	-0.592	0.554

Growth curve analysis using the formula for R function:

$\log(\text{Fixations/Non-fixations}) \sim (\text{Linear} + \text{Quadratic} + \text{Cubic}) * \text{Competitor} + (1) | \text{Subject}$.

Fixed factor: Competitor (Unrelated, Onset). The analysis was performed over the period from 300 to 800 ms after the presentation of the target label onset.

Bold rows indicate significative values.

Cluster-based nonparametric [analysis](#)

Significant clusters are presented in Figure 9. Parameters for the cluster analysis were identical to Experiment 1, except for the time window (from 300 to 1200ms after the target onset). The results showed that participants looked more to the onset phonological competitor than the unrelated one from 300 to 800ms after the target onset ($t_{cluster} = 22.728$, $t_{max} = 5.265$, $p < 0.001$). In contrast, they looked more to the rhyme phonological competitor than the unrelated competitor from 800 to 900 ms after the target onset ($t_{cluster} = 3.689$, $t_{max} = 2.006$, $p = 0.01$).

In sum, both analyses showed that adults participants demonstrate the onset and rhyme phonological effect with a short preview exposure (200 ms) to the array of pictures.

Discussion

Experiment 3 demonstrated that adults display an early phonological onset preference (peaking at around 700 ms) with a 200ms short preview, when the semantic competitor was replaced by a phonological rhyme competitor. Adults' display of the phonological onset preference in the short preview conditions was dependent on whether the phonological onset competitor was the more relevant and stronger match in comparison to the other competitor (the semantic competitor in Experiment 2 or phonological rhyme competitor in Experiment 3). In Experiment 2, the semantic competitor was a temporally stronger competitor than the phonological onset competitor, as the phonological onset competitor ceased to be relevant as soon as the target word had unfolded and suffered from lexical inhibition from the target word. In Experiment 3, despite its diminishing temporal relevance and lexical inhibition from the target word, the phonological onset competitor was still a stronger competitor than

the phonological rhyme competitor (Allopenna et al., 1998; McQueen & Huettig, 2012). Related findings were reported by Apfelbaum et al., (2021) (reviewed in the Introduction) where a phonological onset effect was found in a no-preview condition. The visual stimuli in their study contained a phonological onset competitor to the target object, but no semantic nor visual competitors.

These results demonstrate that the short picture preview time in Experiment 2 was not the primary cause for the absence of any early phonological onset preference (as shown in Apfelbaum et al., 2021). In Experiment 3, an early phonological onset preference was demonstrated despite the reduced time provided to generate phonological codes during picture preview. Instead, the absence of a semantic competitor amongst the pictures permitted phonologically (onset and rhyme) driven matches between the spoken word and the visual objects to become manifest.

General Discussion

This study set out to investigate how visual attention is driven by the matching of visual and auditory linguistic information in adults and toddlers. Specifically, we asked whether a sufficiently long preview time allowing for ‘implicit naming’ is essential to the display of an early phonological onset preference. In Experiment 1, toddlers showed an early phonological onset preference in both the short- and long-preview conditions. In Experiment 2, adults showed an early phonological onset preference in the long-preview condition, but showed no phonological onset preference in the short-preview condition, replicating a pattern of results previously reported by Huettig & McQueen (2007). In Experiment 3, adults showed an early phonological onset preference in a short-preview condition, when the semantic competitor was replaced by a phonological rhyme competitor. In sum, the two key findings of the current study are 1) both toddlers and adults were able to show an early phonological onset preference in short preview conditions, though, 2) adults’ display of an early phonological onset preference in the short preview condition was extinguished by the presence of a semantic competitor. Below we discuss whether our findings are consistent with the phonological-mapping and/or visual-semantic mapping hypotheses, and the developmental difference in language-mediated visual attention in toddlers and adults.

Phonological and Visual-Semantic Mapping

We pitted the phonological mapping hypothesis against the visual-semantic mapping hypothesis by manipulating the length of picture preview time and the related information embedded in visual competitors. The phonological and visual-semantic mapping hypotheses assume different levels of cascaded activation between the auditory and visual representations. As illustrated in *Figure 1*, matching of phonological representations requires extensive processing of the visual stimuli: First, the processing of the picture's (e.g., *bear*) visual form and spatial location, then retrieval of the object's semantic representation and identity, and finally, retrieval of its name and phonological codes. If the phonological code of the auditory label matches or partially matches that of the object, visual attention (eye gaze) can be preferentially directed to that object. Under the phonological mapping hypothesis (*Figure 1a*), a sufficiently extended picture preview time to allow the extraction of the pictures' phonological codes is essential to the display of a phonological onset preference in the short-preview condition because the phonological competitor is only relevant until the target word is recognised. In contrast, according to the visual-semantic mapping hypothesis, cascaded processing of the auditory stimulus has a more important role in driving visual attention (see *Figure 1b*): hearing the onset of the target word (e.g., "*boat*") leads to cascaded activation of its phonological competitors (e.g., *bear*) and the phonological competitor's visual-semantic representations. Under the visual-semantic mapping hypothesis, matching takes places on the visual-semantic level, and the extraction of the pictures' phonological codes prior to target word onset is not essential to the display of a phonological onset preference. Therefore, a short picture preview time should not extinguish the phonological onset preference. Our findings that toddlers and adults were able to display an early phonological onset preference in both short-preview (Experiments 1 and 3) and long-preview (Experiment 1 and 2) conditions appear to be, overall, more consistent with the predictions of the visual-semantic mapping hypothesis. However, the fact that our adult participants performed differently in the two short-preview conditions (Experiments 2 and 3) and long preview condition (Experiment 2) indicate a more complex interpretation involving two interacting factors: picture preview time and the presence of a semantic competitor. The presence of a semantic competitor extinguished the phonological preference only in the short preview condition, but not in the long preview condition, indicating that having a sufficiently

long preview time to identify and implicitly name the pictures leads to a more stable (less susceptible to interference) phonological preference. In contrast, when the semantic competitor is absent, there is no effect of the preview on the phonological onset effect (Apfelbaum et al., 2021 and our Experiment 3).

This leads us to propose the following: a phonological preference in the visual world task can be driven by matching at both the phonological and visual-semantic levels. With a sufficiently long preview time for implicit naming, a match on the phonological level and likely also the visual-semantic levels, leading to a relatively strong and stable phonological preference. With a short preview time where implicit naming is not possible, matching on the visual-semantic levels leads to a relatively weak phonological preference that could be overshadowed by strong top-down factors (e.g., the presence of a semantic competitor, which is a semantic match to the target label). Recent VWP studies provide evidence of top-down semantic pop-out effects in adults (Belke et al., 2008; Moores et al., 2003; Nuthmann et al., 2019). In three VWP experiments with no picture preview (e.g., “*Look for the banana*” followed by a four-picture target-absent display), Nuthmann, De Groot, Huettig, & Olivers (2019) found that adults were more likely to make a first saccade from the central fixation of the screen to a semantic competitor (e.g., *monkey*), in comparison to an unrelated distractor. These findings demonstrate that adults are prone to top-down semantic attentional capture, even when the pictures are presented in the peripheral vision with no prior preview.

Developmental Differences in Exploiting Top-down Contextual Information

Interestingly, toddlers in the short-preview condition did not show diminished preference to the phonological onset competitor in the presence of the semantic competitor. A plausible explanation is that toddlers, in comparison to adults, are less likely to exploit top-down visual-semantic contextual information. A body of VWP research on ambiguous sentence comprehension (e.g., “*Put the frog on the napkin in the box*”) have shown that adults are more likely to exploit the context provided by the visual display to help them comprehend and interpret the linguistic utterances (Spivey et al., 2002; Tanenhaus et al., 1995; Crocker et al., 2010; Knoeferle & Crocker, 2006). Interestingly, 5-year-old children in the same task were driven mainly by the linguistic utterances, and that additional context

provided by the visual display did not change or prompt them to revise their interpretation of the sentence structures (Kidd et al., 2011; Snedeker & Trueswell, 2004; Trueswell et al., 1999).

One piece of evidence that toddlers are more driven by the linguistic utterances in a VWP task is the duration of the phonological onset preference. For adults, the phonological onset preference is typically early and relatively short-lived: preference for the phonological competitor over the unrelated distractor lasts approximately 600 – 800 ms after target word onset (Huetting & McQueen, 2007 and the current study). Thereafter, adults' preference for the phonological competitor becomes similar to or weaker than the unrelated distractor, indicating lexical competition or suppression from the target word (mean duration of the target word is 823 ms in Experiments 1 & 2). For toddlers, the phonological onset preference typically peaks later and lasts longer, approximately 1000 - 1200 ms after target word onset (Chow et al., 2017 and the current study). This pattern of results suggests that toddlers are less able to inhibit phonological competitors, compared to their adult counterparts, and therefore continue to show a preference for the phonological competitor after the target word has fully unfolded.

There is ample evidence that the adult mental lexicon is not only supported by facilitatory connections, but also inhibitory ones. Inhibitory connections allow lexical-semantic items to compete more efficiently for selection, as more plausible word candidates can suppress weaker candidates (McClelland & Elman, 1986). Lexical competition and semantic inhibition effects have been observed in adults and school-aged children using the pause-detection task, picture-naming task and visual world eye-tracking tasks (Dahan et al., 2001; Glaser & Döngelhoff, 1984; Kapnoula et al., 2015; Schriefers et al., 1990; Weighall et al., 2016). However, recent evidence suggests that a lexical-semantic inhibitory mechanism is not initially in-place in the toddler's mental lexicon. Toddler eye-tracking studies using priming paradigms report inhibitory phonological priming effects and inhibitory semantic priming effects emerging between 18- to 24-months of age (Arias-Trejo & Plunkett, 2009, 2013; Chow et al., 2016, 2018; Mani & Plunkett, 2010, 2011; Styles & Plunkett, 2009). These findings indicate that, toddlers are still in the process of constructing inhibitory connections between phono-lexical-semantic representations, and developing adult-like word selection efficiency. They also support the suggestion

that the toddlers in Experiment 1 were less efficient than adults at suppressing the phonological competitors that drive the early phonological onset effects observed in both preview conditions.

Implicit Naming in Short Preview Conditions

An important assumption in the design of the current study and previous studies, is that 100 - 200 ms of preview is insufficient for the extraction of pictures' phonological codes. But how short is too short for the automatic extraction of picture names, i.e., implicit naming? For example, [a recent study](#) have demonstrated that a 0ms preview is sufficient for adults to demonstrate a phonological onset preference in the absence of semantic and/or visual competitors ([Apfelbaum et al., 2021](#)); [in contrast, in school-aged children the phonological onset effect has only been studied with a 500ms preview](#) (e.g., McMurray, Farris-Trimble, & Rigler, 2017; Rigler et al., 2015; Villameriel, Costello, Dias, Giezen, & Carreiras, 2019). It is therefore not possible to completely rule out the phonological mapping hypothesis on the basis of these findings.

Conclusion

To conclude, this study reported that both toddlers and adults were able to show an early phonological onset preference in short preview conditions (100 – 200 ms) of a visual world task. We suggest that the toddlers in the current study were driven more by bottom-up, phonological information when selecting a referent in a visual world task, as compared to adults who were more inclined to exploit top-down, visual-semantic information when directing their attention to a visual object. Our findings take us a step forward into understanding the dynamics and development of the integration of linguistic auditory information and non-linguistic visual information during spoken word comprehension. Future research should investigate how much and what types of information can be extracted from pictures within short preview times, by systematically manipulating the number and type of competitor pictures in the visual display.

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. <https://doi.org/10.1006/jmla.1997.2558>
- Altmann, G. T. M., & Mirković, J. (2009). Incrementality and Prediction in Human Sentence Processing. *Cognitive Science*, 33(4). <https://doi.org/10.1111/j.1551-6709.2009.01022.x>.Incrementality
- Angulo-Chavira, A. Q., & Arias-Trejo, N. (2021). Mediated semantic priming interference in toddlers as seen through pupil dynamics. *Journal of Experimental Child Psychology*, 208(105146), 1–15. <https://doi.org/10.1016/j.jecp.2021.105146>
- Apfelbaum, K. S., Klein-Packard, J., & McMurray, B. (2021). The pictures who shall not be named: Empirical support for benefits of preview in the Visual World Paradigm. *Journal of Memory and Language*, 121, 104279.
- Arias-Trejo, N., Angulo-Chavira, A. Q., Avila-Varela, D. S., Chua-Rodriguez, F., & Mani, N. (2022). Developmental Changes in Phonological and Semantic Priming Effects in Spanish-Speaking Toddlers. *Developmental Psychology*, 58(2), 236–251. <https://doi.org/10.1037/dev0001290>
- Arias-Trejo, N., & Plunkett, K. (2009). Lexical-semantic priming effects during infancy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1536), 3633–3647. <https://doi.org/10.1098/rstb.2009.0146>
- 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. <https://doi.org/10.1016/j.cognition.2013.03.008>
- Belke, E., Humphreys, G. W., Watson, D. G., Meyer, A. S., & Telling, A. L. (2008). Top-down effects of semantic knowledge in visual search are modulated by cognitive but not perceptual load. *Perception and Psychophysics*, 70(8), 1444–1458. <https://doi.org/10.3758/PP.70.8.1444>
- Canfield, R. L., Smith, E. G., Brezsnyak, M. P., & Snow, K. L. (1997). Information processing through the first year of life: a longitudinal study using the visual expectation paradigm. *Monographs of the Society for Research in Child Development*, 62(2), 1–160.

- 1 <https://doi.org/10.2307/1166196>
- 2 Chow, J., Aimola, A., & Plunkett, K. (2017). Spoken-word recognition in 2-year-olds : The tug of war
- 3 between phonological and semantic activation. *Journal of Memory and Language*, 93, 104–134.
- 4 <https://doi.org/10.1016/j.jml.2016.08.004>
- 5 Chow, J., Aimola Davies, A. M., Fuentes, L. J., & Plunkett, K. (2016). Backward Semantic Inhibition
- 6 in Toddlers. *Psychological Science*, 27(10), 1312–1320.
- 7 <https://doi.org/10.1177/0956797616659766>
- 8 Chow, J., Aimola Davies, A. M., Fuentes, L. J., & Plunkett, K. (2018). The vocabulary spurt predicts
- 9 the emergence of backward semantic inhibition in 18-month-old toddlers. *Developmental*
- 10 *Science, September*, e12754. <https://doi.org/10.1111/desc.12754>
- 11 Crocker, M. W., Knoeferle, P., & Mayberry, M. R. (2010). Situated sentence processing: The
- 12 coordinated interplay account and a neurobehavioral model. *Brain and Language*, 112(3), 189–
- 13 201. <https://doi.org/10.1016/j.bandl.2009.03.004>
- 14 Cummings, A., Čeponiene, R., Koyama, A., Saygin, A. P., Townsend, J., & Dick, F. (2006). Auditory
- 15 semantic networks for words and natural sounds. *Brain Research*, 1115(1), 92–107.
- 16 <https://doi.org/10.1016/j.brainres.2006.07.050>
- 17 Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches
- 18 and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive*
- 19 *Processes*, 16(5–6), 507–534. <https://doi.org/10.1080/01690960143000074>
- 20 Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake:
- 21 Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin*
- 22 *and Review*, 12(3), 453–459. <https://doi.org/10.3758/BF03193787>
- 23 Dahan, D., Tanenhaus, M. K., & Salverda, A. P. (2007). The influence of visual processing on
- 24 phonetically driven saccades in the “ visual world ” paradigm. In *Eye Movements: A Window on*
- 25 *Mind and Brain* (Issue April 2018, pp. 471–486). Elsevier. [https://doi.org/10.1016/B978-](https://doi.org/10.1016/B978-008044980-7/50023-9)
- 26 008044980-7/50023-9
- 27 De Groot, F., Huettig, F., & Olivers, C. N. L. (2016). When meaning matters: The temporal dynamics

- 1 of semantic influences on visual attention. *Journal of Experimental Psychology: Human*
- 2 *Perception and Performance*, 42(2), 180–196. <https://doi.org/10.1037/xhp0000102>
- 3 Dink, J. W., & Ferguson, B. (2015). *eyetrackingR: An R library for eye-tracking data analysis*.
- 4 Duñabeitia, J. A., Avilés, A., Afonso, O., Scheepers, C., & Carreiras, M. (2009). Qualitative
- 5 differences in the representation of abstract versus concrete words: Evidence from the visual-
- 6 world paradigm. *Cognition*, 110(2), 284–292. <https://doi.org/10.1016/j.cognition.2008.11.012>
- 7 Duta, M. D., Styles, S. J., & Plunkett, K. (2012). ERP correlates of unexpected word forms in a
- 8 picture-word study of infants and adults. *Developmental Cognitive Neuroscience*, 2(2), 223–234.
- 9 <https://doi.org/10.1016/j.dcn.2012.01.003>
- 10 Ferreira, F., & Tanenhaus, M. K. (2007). Introduction to the special issue on language-vision
- 11 interactions. *Journal of Memory and Language*, 57(4), 455–459.
- 12 <https://doi.org/10.1016/j.jml.2007.08.002>
- 13 Glaser, W. R., & Döngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of*
- 14 *Experimental Psychology: Human Perception and Performance*, 10(5), 640–654.
- 15 <https://doi.org/10.1037/0096-1523.10.5.640>
- 16 Green, P., & Macleod, C. J. (2016). SIMR: An R package for power analysis of generalized linear
- 17 mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498.
- 18 <https://doi.org/10.1111/2041-210X.12504>
- 19 Hamilton, a, Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a
- 20 British communicative development inventory. *Journal of Child Language*, 27(3), 689–705.
- 21 <https://doi.org/10.1017/S0305000900004414>
- 22 Huettig, F, & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated
- 23 attention is based on lexical input and not modulated by contextual appropriateness. *Visual*
- 24 *Cognition*, 15(8), 985–1018. <https://doi.org/10.1080/13506280601130875>
- 25 Huettig, F, & McQueen, J. M. (2007). The tug of war between phonological , semantic and shape
- 26 information in language-mediated visual search. *Journal of Memory and Language*, 57, 460–
- 27 482. <https://doi.org/10.1016/j.jml.2007.02.001>

- 1 Huettig, Falk, & Altmann, G. T. M. (2004). The on-line processing of ambiguous and unambiguous
2 words in context: Evidence from head-mounted eyetracking. *The On-Line Study of Sentence*
3 *Comprehension: Eyetracking, ERPs and Beyond*, 187–207.
4 <https://doi.org/10.4324/9780203509050>
- 5 Huettig, Falk, & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic
6 competitor effects and the visual world paradigm. *Cognition*, 96(1), 23–32.
7 <https://doi.org/10.1016/j.cognition.2004.10.003>
- 8 Huettig, Falk, & Altmann, G. T. M. (2011). Looking at anything that is green when hearing “frog”:
9 How object surface colour and stored object colour knowledge influence language-mediated
10 overt attention. *Quarterly Journal of Experimental Psychology*, 64(1), 122–145.
11 <https://doi.org/10.1080/17470218.2010.481474>
- 12 Huettig, Falk, Mishra, R. K., & Olivers, C. N. L. (2012). Mechanisms and representations of
13 language-mediated visual attention. *Frontiers in Psychology*, 3(JAN), 1–11.
14 <https://doi.org/10.3389/fpsyg.2011.00394>
- 15 Johnson, E. K., McQueen, J. M., & Huettig, F. (2011). Toddlers’ language-mediated visual search:
16 They need not have the words for it. *Quarterly Journal of Experimental Psychology* (2006),
17 64(9), 1672–1682.
- 18 Kapnoula, E. C., Packard, S., Gupta, P., & McMurray, B. (2015). Immediate lexical integration of
19 novel word forms. *Cognition*, 134, 85–99. <https://doi.org/10.1016/j.cognition.2014.09.007>
- 20 Kidd, E., Stewart, A. J., & Serratrice, L. (2011). Children do not overcome lexical biases where adults
21 do: The role of the referential scene in garden-path recovery. *Journal of Child Language*, 38(1),
22 222–234. <https://doi.org/10.1017/S0305000909990316>
- 23 Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world
24 knowledge: Evidence from eye tracking. *Cognitive Science*, 30(3), 481–529.
25 https://doi.org/10.1207/s15516709cog0000_65
- 26 Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the
27 immediate visual context on incremental thematic role-assignment: Evidence from eye-

- 1 movements in depicted events. *Cognition*, 95(1), 95–127.
- 2 <https://doi.org/10.1016/j.cognition.2004.03.002>
- 3 Levelt, W. J. M., Roelofs, A. P. ., & Meyer, A. . (1999). A theory of lexical access in speech
- 4 production [target paper]. *Behavioral and Brain Sciences*, 22(1999), 1–37.
- 5 Mani, N., Johnson, E., McQueen, J. M., & Huettig, F. (2012). How Yellow Is Your Banana?
- 6 Toddlers' Language-Mediated Visual Search in Referent-Present Tasks. *Developmental*
- 7 *Psychology*, 1–31.
- 8 Mani, N., & Plunkett, K. (2010). In the infant's mind's ear: evidence for implicit naming in 18-
- 9 month-olds. *Psychological Science*, 21(7), 908–913. <https://doi.org/10.1177/0956797610373371>
- 10 Mani, N., & Plunkett, K. (2011). Phonological priming and cohort effects in toddlers. *Cognition*,
- 11 121(2), 196–206. <https://doi.org/10.1016/j.cognition.2011.06.013>
- 12 Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal*
- 13 *of Neuroscience Methods*, 164, 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- 14 McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive*
- 15 *Psychology*, 18(1), 1–86. [https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)
- 16 McMurray, B., Farris-trimble, A., & Rigler, H. (2017). Waiting for lexical access : Cochlear implants
- 17 or severely degraded input lead listeners to process speech less incrementally. *Cognition*,
- 18 169(September), 147–164. <https://doi.org/10.1016/j.cognition.2017.08.013>
- 19 McQueen, J. M., & Huettig, F. (2012). Changing only the probability that spoken words will be
- 20 distorted changes how they are recognized. *The Journal of the Acoustical Society of America*,
- 21 131(1), 509–517. <https://doi.org/10.1121/1.3664087>
- 22 McQueen, J. M., & Huettig, F. (2014). Interference of spoken word recognition through phonological
- 23 priming from visual objects and printed words. *Attention, Perception, and Psychophysics*, 76(1),
- 24 190–200. <https://doi.org/10.3758/s13414-013-0560-8>
- 25 Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names
- 26 in visual search. *Psychonomic Bulletin and Review*, 14(4), 710–716.
- 27 <https://doi.org/10.3758/BF03196826>

- 1 Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture
2 interference paradigm. *Memory and Cognition*, 35(3), 494–503.
3 <https://doi.org/10.3758/BF03193289>
- 4 Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual
5 selective attention. *Nature Neuroscience*, 6(2), 182–189. <https://doi.org/10.1038/nn996>
- 6 Morsella, E., & Miozzo, M. (2002). Evidence for a Cascade Model of Lexical Access in Speech
7 Production. *Journal of Experimental Psychology: Learning Memory and Cognition*, 28(3), 555–
8 563. <https://doi.org/10.1037/0278-7393.28.3.555>
- 9 Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a
10 cascade model of lexical access. *Journal of Memory and Language*, 53(3), 359–377.
11 <https://doi.org/10.1016/j.jml.2005.05.001>
- 12 Nuthmann, A., De Groot, F., Huettig, F., & Olivers, C. N. L. (2019). Extrafoveal attentional capture
13 by object semantics. *PLoS ONE*, 14(5), 1–19. <https://doi.org/10.1371/journal.pone.0217051>
- 14 Potter, M. C. (1975). Meaning in visual search. *Science*, 187(March), 965–966.
- 15 Rigler, H., Farris-trimble, A., Greiner, L., Walker, J., Tomblin, J. B., & McMurray, B. (2015). *The*
16 *Slow Developmental Time Course of Real-Time Spoken Word Recognition*. 51(12), 1690–1703.
- 17 Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in
18 language production: Picture-word interference studies. *Journal of Memory and Language*,
19 29(1), 86–102. [https://doi.org/10.1016/0749-596X\(90\)90011-N](https://doi.org/10.1016/0749-596X(90)90011-N)
- 20 Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental
21 semantic interpretation through contextual representation. *Cognition*, 71(2), 109–147.
22 [https://doi.org/10.1016/S0010-0277\(99\)00025-6](https://doi.org/10.1016/S0010-0277(99)00025-6)
- 23 Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of
24 lexical-biases and referential scenes in child and adult sentence processing. *Cognitive*
25 *Psychology*, 49(3), 238–299. <https://doi.org/10.1016/j.cogpsych.2004.03.001>
- 26 Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and
27 spoken language comprehension: Effects of visual context on syntactic ambiguity resolution.

- Cognitive Psychology*, 45(4), 447–481. [https://doi.org/10.1016/S0010-0285\(02\)00503-0](https://doi.org/10.1016/S0010-0285(02)00503-0)
- 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. <https://doi.org/10.1016/j.jml.2011.06.002>
- Styles, S. J., & Plunkett, K. (2009). How do infants build a semantic system? *Language and Cognition*, 1(1), 1–24. <https://doi.org/10.1515/langcog.2009.001>
- Styles, S. J., Plunkett, K., & Duta, M. D. (2015). Infant VEPs reveal neural correlates of implicit naming: Lateralized differences between lexicalized versus name-unknown pictures. *Neuropsychologia*, 77, 177–184. <https://doi.org/10.1016/j.neuropsychologia.2015.07.027>
- Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, 29(6), 557–580. <https://doi.org/10.1023/A:1026464108329>
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634. <https://doi.org/10.1126/science.7777863>
- Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73(2), 89–134. [https://doi.org/10.1016/S0010-0277\(99\)00032-3](https://doi.org/10.1016/S0010-0277(99)00032-3)
- Villameriel, S., Costello, B., Dias, P., Giezen, M., & Carreiras, M. (2019). Language modality shapes the dynamics of word and sign recognition. *Cognition*, 191(May 2018), 103979. <https://doi.org/10.1016/j.cognition.2019.05.016>
- Weighall, A. R., Henderson, L. M., Barr, D. J., Cairney, S. A., & Gaskell, M. G. (2016). Eye-tracking the time-course of novel word learning and lexical competition in adults and children. *Brain and Language*. <https://doi.org/10.1016/j.bandl.2016.07.010>
- Yee, E., Huffstetler, S., & Thompson-Schill, S. L. (2011). Function follows form: Activation of shape and function features during object identification. *Journal of Experimental Psychology*:

- 1 General, 140(3), 348–363. <https://doi.org/10.1037/a0022840>
- 2 Yee, E., Overton, E., & Thompson-Schill, S. L. (2009). Looking for Meaning: Eye Movements Are
- 3 Sensitive to Overlapping Semantic Features, Not Association. *Psychonomic Bulletin & Review*,
- 4 16(5), 869–874. <https://doi.org/10.3758/PBR.16.5.869>. Looking
- 5 Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation
- 6 during spoken word recognition. *Journal of Experimental Psychology: Learning Memory and*
- 7 *Cognition*, 32(1), 1–14. <https://doi.org/10.1037/0278-7393.32.1.1>
- 8
- 9

Appendix A

Table A1. Experiments 1 and 2 test trial target words and the yoked-quadruplets of picture stimuli. The 24 pictures were grouped into six sets of four pictures. Each quadruple of pictures was yoked four times (within-participant) so that, with a different target word (e.g. boat, trousers, frog and Hoover respectively), each picture served once as the phonological onset competitor, semantic competitor, unrelated distractor 1 and unrelated distractor 2

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

1 Table A2. Example of Trial Order in Experiments 1 and 2.

Design Block	Trial Order	Trial Type	Target Word		Pictures (Picture Role)						
0	1	Filler	Juice	Juice	(target)	Nappy	(unrelated)	Brush	(unrelated)	Keys	(unrelated)
1	2	Filler	Fire-engine	Fire-engine	(target)	Fish	(onset)	Aeroplane	(semantic)	Grape	(unrelated)
	3	Test	Trousers	Train	(onset)	Hat	(semantic)	Fridge	(unrelated)	Bear	(unrelated)
	4	Test	Squirrel	Soap	(onset)	Duck	(semantic)	Blocks	(unrelated)	Tomato	(unrelated)
	5	Filler	Chocolate	Chocolate	(target)	Chicken	(onset)	Icecream	(semantic)	Umbrella	(unrelated)
	6	Test	Bath	Bicycle	(onset)	Sofa	(semantic)	Carrot	(unrelated)	Turtle	(unrelated)
	7	Test	Toast	TV	(onset)	Pear	(semantic)	Horse	(unrelated)	Car	(unrelated)
	8	Filler	Bib	Bib	(target)	Book	(onset)	Shoes	(semantic)	Clock	(unrelated)
	9	Test	Digger	Dress	(onset)	Bus	(semantic)	Cat	(unrelated)	Sandwich	(unrelated)
	10	Test	Boots	Butterfly	(onset)	Coat	(semantic)	Peas	(unrelated)	Tractor	(unrelated)
2	11	Filler	Monkey	Monkey	(target)	Milk	(onset)	Dog	(semantic)	Telephone	(unrelated)
	12	Test	Boat	Bear	(onset)	Train	(semantic)	Hat	(unrelated)	Fridge	(unrelated)
	13	Test	Cow	Carrot	(onset)	Turtle	(semantic)	Bicycle	(unrelated)	Sofa	(unrelated)
	14	Filler	Spoon	Spoon	(target)	Swing	(onset)	Glass	(semantic)	Ambulance	(unrelated)
	15	Test	Tiger	Tractor	(onset)	Butterfly	(semantic)	Coat	(unrelated)	Peas	(unrelated)
	16	Test	Cucumber	Cat	(onset)	Sandwich	(semantic)	Dress	(unrelated)	Bus	(unrelated)
	17	Filler	Teddy	Teddy	(target)	Table	(onset)	Ball	(semantic)	Flower	(unrelated)
	18	Test	Helicopter	Horse	(onset)	Car	(semantic)	TV	(unrelated)	Pear	(unrelated)
	19	Test	Doll	Duck	(onset)	Blocks	(semantic)	Tomato	(unrelated)	Soap	(unrelated)
3	20	Filler	Pig	Pig	(target)	Pen	(onset)	Elephant	(semantic)	Button	(unrelated)
	21	Test	Strawberry	Sofa	(onset)	Carrot	(semantic)	Turtle	(unrelated)	Bicycle	(unrelated)
	22	Test	Bee	Bus	(onset)	Cat	(semantic)	Sandwich	(unrelated)	Dress	(unrelated)
	23	Filler	Hand	Hand	(target)	Raisins	(unrelated)	Tree	(unrelated)	Window	(unrelated)
	24	Test	Cake	Coat	(onset)	Peas	(semantic)	Tractor	(unrelated)	Butterfly	(unrelated)
	25	Test	Hoover	Hat	(onset)	Fridge	(semantic)	Bear	(unrelated)	Train	(unrelated)
	26	Filler	Cup	Cup	(target)	Caterpillar	(onset)	Fork	(semantic)	Watch	(unrelated)
	27	Test	Towel	Tomato	(onset)	Soap	(semantic)	Duck	(unrelated)	Blocks	(unrelated)
	28	Test	Cot	Car	(onset)	TV	(semantic)	Pear	(unrelated)	Horse	(unrelated)
4	29	Filler	Spade	Spade	(target)	Snail	(onset)	Watering Can	(semantic)	Door	(unrelated)
	30	Test	Pushchair	Peas	(onset)	Tractor	(semantic)	Butterfly	(unrelated)	Coat	(unrelated)
	31	Test	Sock	Sandwich	(onset)	Dress	(semantic)	Bus	(unrelated)	Cat	(unrelated)
	32	Filler	Bunny	Bunny	(target)	Bag	(onset)	Lion	(semantic)	Plate	(unrelated)
	33	Test	Truck	Turtle	(onset)	Bicycle	(semantic)	Sofa	(unrelated)	Carrot	(unrelated)
	34	Test	Frog	Fridge	(onset)	Bear	(semantic)	Train	(unrelated)	Hat	(unrelated)
	35	Filler	Slide	Slide	(target)	Cooker	(unrelated)	Orange	(unrelated)	Toothbrush	(unrelated)
	36	Test	Banana	Blocks	(onset)	Tomato	(semantic)	Soap	(unrelated)	Duck	(unrelated)
	37	Test	Penguin	Pear	(onset)	Horse	(semantic)	Car	(unrelated)	TV	(unrelated)

2

3

Table A3. Example of Trial Order in Experiment 3.

Design Block	Trial Order	Trial Type	Target Word		Pictures (Picture Role)						
0	1	filler	boot	boot	(target)	mouth	(unrelated)	stairs	(unrelated)	key	(unrelated)
	2	filler	glass	glass	(target)	bird	(unrelated)	slide	(unrelated)	tongue	(unrelated)
1	3	filler	owl	owl	(target)	bread	(unrelated)	ear	(unrelated)	stone	(unrelated)
	4	test	star	sock	(onset)	car	(rhyme)	door	(unrelated)	pen	(unrelated)
	5	filler	bed	bed	(target)	lamp	(unrelated)	nose	(unrelated)	milk	(unrelated)
	6	test	house	hand	(onset)	mouse	(rhyme)	spoon	(unrelated)	cot	(unrelated)
	7	filler	cake	cake	(target)	sheep	(unrelated)	bowl	(unrelated)	arm	(unrelated)
2	8	filler	plate	plate	(target)	cow	(unrelated)	swing	(unrelated)	bread	(unrelated)
	9	test	frog	foot	(onset)	dog	(rhyme)	truck	(unrelated)	boat	(unrelated)
	10	filler	book	book	(target)	shoe	(unrelated)	plane	(unrelated)	ear	(unrelated)
	11	test	cat	cup	(onset)	hat	(rhyme)	chair	(unrelated)	door	(unrelated)
	12	filler	pig	pig	(target)	box	(unrelated)	train	(unrelated)	eye	(unrelated)
3	13	filler	fork	fork	(target)	horse	(unrelated)	brush	(unrelated)	stone	(unrelated)
	14	test	trees	truck	(onset)	cheese	(rhyme)	bus	(unrelated)	cot	(unrelated)
	15	filler	tooth	tooth	(target)	dress	(unrelated)	bird	(unrelated)	stairs	(unrelated)
	16	test	hair	hat	(onset)	chair	(rhyme)	mouse	(unrelated)	cup	(unrelated)
	17	filler	peas	peas	(target)	watch	(unrelated)	sheep	(unrelated)	box	(unrelated)
4	18	filler	blocks	blocks	(target)	mouth	(unrelated)	lamp	(unrelated)	cow	(unrelated)
	19	test	boat	bear	(onset)	coat	(rhyme)	cheese	(unrelated)	foot	(unrelated)
	20	filler	juice	juice	(target)	fish	(unrelated)	stairs	(unrelated)	arm	(unrelated)
	21	test	pup	pear	(onset)	cup	(rhyme)	bus	(unrelated)	sock	(unrelated)
	22	filler	doll	doll	(target)	bird	(unrelated)	nose	(unrelated)	swing	(unrelated)
5	23	filler	toast	toast	(target)	horse	(unrelated)	slide	(unrelated)	eye	(unrelated)
	24	test	moon	mouse	(onset)	spoon	(rhyme)	hand	(unrelated)	ship	(unrelated)
	25	filler	soap	soap	(target)	lamp	(unrelated)	fish	(unrelated)	train	(unrelated)
	26	test	clock	coat	(onset)	sock	(rhyme)	bear	(unrelated)	pen	(unrelated)
	27	filler	toes	toes	(target)	plane	(unrelated)	bread	(unrelated)	key	(unrelated)
6	28	filler	egg	egg	(target)	cow	(unrelated)	nose	(unrelated)	bowl	(unrelated)
	29	test	chip	chair	(onset)	ship	(rhyme)	foot	(unrelated)	dog	(unrelated)
	30	filler	bag	bag	(target)	milk	(unrelated)	tongue	(unrelated)	stone	(unrelated)
	31	test	bear	boat	(onset)	pear	(rhyme)	car	(unrelated)	shoes	(unrelated)
	32	filler	sink	sink	(target)	mouth	(unrelated)	box	(unrelated)	horse	(unrelated)
7	33	filler	leg	leg	(target)	slide	(unrelated)	brush	(unrelated)	milk	(unrelated)
	34	test	duck	dog	(onset)	truck	(rhyme)	coat	(unrelated)	ship	(unrelated)
	35	filler	shirt	shirt	(target)	bowl	(unrelated)	train	(unrelated)	ear	(unrelated)
	36	test	sand	spoon	(onset)	hand	(rhyme)	bus	(unrelated)	shoes	(unrelated)
	37	filler	bee	bee	(target)	swing	(unrelated)	plane	(unrelated)	eye	(unrelated)
8	38	filler	ball	ball	(target)	key	(unrelated)	arm	(unrelated)	sheep	(unrelated)
	39	test	coat	car	(onset)	boat	(rhyme)	pear	(unrelated)	door	(unrelated)
	40	filler	bath	bath	(target)	tongue	(unrelated)	dress	(unrelated)	watch	(unrelated)
	41	test	chair	cheese	(onset)	bear	(rhyme)	hat	(unrelated)	pen	(unrelated)

Author's note

J.C. designed and collected data for the experiment, processed and analysed the data, and wrote the manuscript; [A.A. analysed the data and wrote the manuscript](#); M.S. designed, collected, processed and analysed data for the experiment; L.H. collected, processed and analysed data for the experiment; K.P designed the experiment, and wrote and critically reviewed the manuscript.

The Authors declare that there is no conflict of interest.

J.C. is funded by the Winkler Career Development Fellowship at St Hugh's College, Oxford.

Correspondence concerning this article should be addressed to Kim Plunkett, Department of Experimental Psychology, Anna Watts Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, U.K. Telephone: +44 (0)1865 271384. Email: kim.plunkett@psy.ox.ac.uk

We thank Amanda Griffins and Irina Lepadatu for participant recruitment and data collection, and all the adult, toddler and parent participants who took part in the study.

The data and analysis codes that support this research are openly available in https://osf.io/uchdv/?view_only=6ee0f0aa68c44a08b90c9dbc1fd1c195