HARVARD
Kenneth C. Griffin

GRADUATE SCHOOL
OF ARTS AND SCIENCES

THESIS ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Psychology

have examined a dissertation entitled

Go with the flow: Investigating the dynamics of lexical access in child language production and comprehension

presented by Margaret C. Kandel

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature _____

Typed name: Prof. Jesse Snedeker

Signature _____

Typed name: Prof. Elika Bergelson

Signature _____

Typed name: Prof. Kathryn Davidson

Date: April 17, 2025

Go with the flow:

Investigating the dynamics of lexical access in child language production and comprehension

A dissertation presented

by

Margaret C. Kandel

to

The Department of Psychology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Psychology

Harvard University

Cambridge, Massachusetts

April 2025

Dissertation Advisor: Jesse Snedeker Margaret C. Kandel

Go with the flow:

Investigating the dynamics of lexical access in child language production and comprehension

Abstract

Producing and comprehending language is no small feat, involving many different steps and types of representations. Despite this complexity, the language system is employed seemingly effortlessly by adults, and acquiring this system comes naturally to children. In adults, rapid and efficient language processing is enabled by the incremental, cascading, and interactive transfer of information across levels of representation. In this dissertation, I explore whether these interactive informational cascades are already present in the language system by early childhood or whether they develop later (e.g., with increased language experience or as the brain matures), and I probe the limitations of webcam eye-tracking approaches to ask these questions. I present three papers investigating the dynamics of information flow during four to five-year-old children's language processing, using lexical processing as a case study. Paper 1 used a picture naming task to test for informational cascades during word production, finding evidence that cascaded processing arises in the language system by at least five years of age. Paper 2 used visual world eye-tracking to test whether informational cascades in four and five-year-old children are interactive, showing that young children are able to use top-down contextual information during language comprehension to pre-activate upcoming word representations and constrain bottom-up processing. Paper 3 assessed the suitability of webcam eye-tracking

techniques for studying moment-to-moment processing in young children by testing the sensitivity of two webcam eye-tracking methods to detect evidence of incrementality in children's word recognition. Taken together, the results of these studies demonstrate that the incremental, interactive informational cascades that are integral to adult language processing are robust and active early in life, suggesting that they are fundamental consequences of the language system architecture.

Table of Contents

Acknowledgements

Thank you to the many incredible people who have contributed to and supported this work. First and foremost, I would like to thank Jesse Snedeker for advising and guiding me throughout my time in graduate school and for making me into the scientist I am today. I would also like to thank my additional committee members Elika Bergelson and Kathryn Davidson for challenging me to further refine my ideas and to think about my projects in different lights. Thank you to my co-authors and to the army of lab managers and research assistants who supported this work, including Durgesh Rajandiran, Briony Waite, Hanna Shine, Amelia Harris, Antara Bhattacharya, Beatriz Leitão, Benazir Neree, Briony Waite, Danielle Novak, Eleanor Muir, Emily Liu, Hannah Kleiner, Hanna Shine, Iandra Ramos, Joanna Lau, Mikaela Martin, Madeleine Presgrave, Parker Robbins, and Timothy Guest. Thank you as well to Alfonso Caramazza, Anthony Yacovone, Joshua Cetron, Mieke Slim, and Patrick Mair for sharing their thoughts on these projects, and thank you to all of the individuals and families who participated in the experiments.

I would also like to thank the additional mentors I have had along my research journey. Thank you to Raffaella Zanuttini for welcoming me to the world of linguistics, and thank you to Colin Phillips for introducing me to the world of psycholinguistics; I would not be where I am today if you had not taken a chance on me. Thank you as well to the other members of the Snedeker Lab, the Harvard Laboratory for Developmental Studies group, and my graduate year cohort for showing me the ropes and for providing support and camaraderie within the lab and beyond.

Finally, thank you to my family, my partner, and my friends for all of the various ways you have supported me over the years. I am so lucky to have you all by my side.

Chapter 1

[Introduction]

1.1. Introduction

Language allows humans to convey complex meanings in the form of arbitrary symbols, such as sequences of sounds (in the case of spoken language) or gestures (in the case of signed language). Language plays an integral role in everyday life, with adults producing approximately 16,000 spoken words per day (Mehl et al., 2007) and encountering many more in both spoken and written form. In most cases, the language system is employed seemingly effortlessly, recalling the old adage "*it's so easy, a child could do it*" — and in fact, they do. Children begin mapping words to meanings as young as six months of age (Bergelson & Aslin, 2017; Bergelson & Swingley, 2012, 2013, 2015; Tincoff & Juscyzk, 1999, 2011), begin stringing words together in multiword utterances by the second year of life (Miller & Chapman, 1981), and become prolific producers of sentences by the age of three (Rice et al., 2010). This language learning appears to come naturally to children; virtually all children exposed to a first language become proficient users of that language.

Nevertheless, producing and comprehending language is no small feat, even for adults. Assembling and decoding the linguistic signals that map between meaning and form are complex processes that involve many different steps and types of representations. For instance, in language production, starting with a concept or abstract message that they want to convey, speakers must select the words from their lexicon that best express the ideas within this concept

(for an adult, this active working lexicon may contain tens of thousands of competing entries to choose from; see Goulden et al., 1990; Nagy et al., 1985; Zareva et al., 2005 for estimates), choose an appropriate syntactic structure to convey the message, arrange the words of the message in a grammatical sequence, change word forms as necessary (e.g. verb inflection), assemble the sequence of words into prosodic units, and execute a motor plan to articulate these units. As speakers formulate their utterances, they may additionally take into account constraints such as their knowledge of the common ground, the social context, or their expectations of the interlocutor's language ability or background knowledge (Andersen, 2014; Clark & Brennan, 1991; Holler & Welkin, 2009; Horton & Keysar, 1996; Krauss & Weinheimer, 1966; Lee & Pinker, 2010; Nilsen & Graham, 2009; Sachs & Devin, 1976). On the comprehension side, interlocutors must parse the linguistic units that they encounter in the stream of language input, identify the words they are hearing, assemble them together in a meaningful structure, and interpret the intended message of that structure. All of this takes place through channels that are noisy (subject to background noise and speech distortions), with input that can contain ambiguity at both the word level (e.g., homophones) and the structural level (e.g., prepositional phrase attachment ambiguity). Yet, despite the complexity of language processing, information moves through the language system at a rapid rate, with adult users producing and able to interpret speech rates of two to three words per second (approximately 5.65 syllables per second) in natural conversation, with bursts of up to seven words per second (Bock, 1995; Deese, 1984; Maclay & Osgood, 1959).

In this dissertation, I take a developmental approach to explore three key features of the language system that enable this rapid processing: informational cascades (Chapter 2), interactivity between levels of representation (Chapter 3), and incrementality (in particular, the

methods that can be used to measure incrementality and other features of real-time processing in young children; Chapter 4). This work provides insight into how and when the dynamic processing of the language system develops as well as the methods that can be used to assess moment-to-moment processing in children. As a case study, I focus on language processing at the word level — i.e., *lexical processing*. I explore how lexical representations are accessed for language production and comprehension in early childhood (4–5 years of age) and whether lexical processing in the developing language system resembles that of adults. Understanding when features of language processing appear in the developing language system provides insight into whether those features are fundamental consequences of the mind's architecture or whether they develop later in life as a by-product of experience. Examining the early stages of the language system is important for constructing models of language processing and cognition that apply across all stages of life.

The remainder of this Introduction is organized as follows. First, I provide an overview of the structure of the mature language system and how information flows through it in language production and comprehension.[1] Second, I explain the theoretical reasons why one might expect to see differences in information flow within young children's language processing. Third, I introduce eye-tracking as a method to probe children's moment-to-moment language comprehension and discuss the potential benefits and challenges of moving this method to a web-based setting (the focus of Chapter 4). Finally, I present the goals of the dissertation and describe the comprising chapters.

[1] In this dissertation, I refer to a single language system that comprises both language production and comprehension processes. Under such a hypothesized unified system, language production and comprehension utilize a shared set of cognitive mechanisms and representations, with different inputs and directions of informational flow leading to different outputs (e.g., Dell & Chang, 2014; Lewis & Phillips, 2015; Pickering & Garrod, 2013b). However, the conclusions of the research within this dissertation are not contingent upon this assumption.

1.1.1. The structure of the mature language system

Language production and comprehension involve multiple distinct, interconnected levels of processing, each associated with different representations (e.g., Butterworth, 1989; Cutler & Clifton, 1999; Dell, 1986; Garrett, 1980; Kuperberg & Jaeger, 2016; Levelt, 1989; inter alia). These levels are described below, along with how they relate to lexical processing.

Contemporary models of language production tend to converge on three major processing levels prior to articulation: the message level, the grammatical encoding level, and the phonological encoding level (e.g., Bock & Levelt, 1994). Messages are conceptual structures that represent the intended meaning to be conveyed in an utterance. At the message level, speakers conceptualize what it is that they would like to express, and then this message serves as input for grammatical encoding. Grammatical encoding is the process of formulating the message in a linguistic format. This level includes both functional processing and positional processing: Functional processing consists of lexical selection (identifying the lexical items that will best convey the message) and function assignment (assigning these lexical elements syntactic functions and relations; e.g., subject–nominative). Positional processing consists of constituent assembly (assembling the selected lexical elements into a syntactic structure) and inflection. The phonological encoding level prepares the utterance for articulation by retrieving and assembling the morpho-phonological forms of the words in the utterance and dividing the utterance into prosodic units. After phonological encoding, the speaker engages articulatory planning to produce the utterance.

The primary challenges for lexical processing in language production are: (i) determining the lexical concept (e.g., Levelt et al., 1999) to be conveyed (conceptual/message-level

processing), (ii) identifying the best-suited lexical representation from the mental lexicon to convey the lexical concept (lexical selection), and (iii) accessing the phonological form of the word to be uttered (phonological encoding). The mental representations involved in these processes are often modelled as networks of interconnected nodes; in this framework, accessing a representation is modelled as activating its corresponding node (e.g., Bloem & La Heij, 2003; Caramazza, 1997; Dell, 1986; Dell et al., 1997; Levelt et al., 1999; Roelofs, 1992). Within network models of lexical processing, activation is hypothesized to spread along the connections that link nodes, both across and between representation types (e.g., Collins & Loftus, 1975; Dell, 1986; Roelofs, 1992; Dell et al., 1997; Levelt et al., 1999). For instance, activated conceptual nodes spread their activation to connected, related lexical nodes, which in turn spread activation to phonological nodes, allowing speakers to access the lexical representations required to convey the intended concept.

Similar to language production, contemporary models of language comprehension divide the comprehension process into distinct levels of processing (see, e.g., Snedeker, 2009 for a schematic break-down of these processes). The first levels involve perceptual processing of the linguistic input (e.g., sounds or gestures), segmenting this input into discrete units (phonological processing), and processing prosodic cues (prosodic processing). The processed input is mapped onto stored representations in the comprehender's mental lexicon during word recognition, and these word representations are retrieved during lexical access. Comprehenders use prosodic and lexical information to identify the structure of the sentence they hear during syntactic parsing. Comprehenders then interpret lexical and syntactic representations (semantic analysis) and integrate this information with the discourse or broader context (pragmatic processing) to build a representation of the intended message.

The primary challenges for lexical processing in language comprehension are thus: (i) parsing the word from the input (phonological processing), (ii) retrieving the word representation and its meaning (lexical access), and (iii) integrating the word into the sentence and discourse. As in language production, lexical processing in language comprehension can be modeled as activating phonological, lexical, and conceptual representations.

A central question in psycholinguistic research is how information is passed between processing levels in both language production and comprehension. The flow of information from lower levels of representation (those closer to articulatory or perceptual processes) to higher levels (those representing complex meanings) is referred to as bottom-up processing. The flow of information in the opposing direction is referred to as top-down processing. Language production can thus be viewed as a primarily top-down process, in which speakers start from a conceptual message-level representation and convert it into a series of phonological representations to be articulated. In contrast, language comprehension can be viewed as a largely bottom-up process, in which speakers start with phonological representations identified from the input and build a representation of the conceptual message intended by the speaker. However, in adult language processing, information does not strictly flow in a single direction in either production or comprehension. The next section describes three key features of information flow in the mature language system: incrementality, informational cascades, and interactivity.

1.1.2. Adult language processing is incremental, cascading, and interactive

It is widely accepted that during language processing, information flows in an incremental and cascaded fashion — meaning that information passes between levels of processing before processing at one level is complete. For instance, in language production,

speakers do not fully plan a full utterance before speech onset (see e.g., Bock, 1982; Bock & Levelt, 1994; Kempen & Hoenkamp, 1987; Levelt, 1989), and there is evidence that during lexical processing, speakers begin accessing the phonological forms of words under consideration for production even before selecting which word to incorporate into their utterance (e.g., Costa et al., 2000; Cutting & Ferreira, 1999; Jescheniak & Schriefers, 1998; Morsella & Miozzo, 2002; Peterson & Savoy, 1998; Rapp & Goldrick, 2000; Starreveld & La Heij, 1995; see Chapter 2 for detail). In spoken language comprehension, comprehenders do not wait until they have heard a full utterance to begin constructing a potential syntactic and semantic parse (leading, e.g., to garden path effects; Ferreira & Clifton, 1986; Garnsey et al., 1997; inter alia), and during lexical selection, listeners map the acoustic-phonetic input to candidate words in their mental lexicon incrementally, rather than waiting until they have heard an entire word before activating lexical representations (e.g., Allopenna et al., 1998; see Chapters 3 and 4).

There is also evidence that the language system is interactive across levels of representation and processing such that processing at one level can be influenced by information from levels both above it (top-down influences) and below it (bottom-up influences). In language production, this interactivity means that a word selected during lexical access is not only informed by the meaning the speaker intends to convey but can also be influenced by activated phonological forms (e.g., in feedback loops; Dell, 1986; Dell & O'Seaghdha, 1991, 1992; Dell et al., 1997; Harley, 1993). In language comprehension, information about likely sentence meanings, the syntactic context, and the speaker's intentions can influence the processing of the words that appear in the input (e.g., via pre-activation of word forms and/or meanings; Ito, 2024; Kuperberg & Jaeger, 2016; Ryskin & Nieuwland, 2023; see Chapter 3). Moreover, there is evidence that information flows through the levels of the language system in a flexible (and

perhaps even strategic) fashion — for instance, how far in advance speakers plan words during language production is modulated by factors such as available planning time (e.g., Ferreira & Swets; Griffin, 2003) and cognitive load (e.g., Wagner et al., 2010), and the top-down processes engaged during language comprehension appear to depend on how predictable the input is given the context (e.g., Brothers et al., 2019; Lau et al., 2013; Lau et al., 2016).

These three features of the language system — incrementality, informational cascades, and interactivity — have functional advantages for language processing. In language production, planning and articulating utterances in incremental chunks distributes cognitive processing over time and reduces the amount of linguistic information stored in working memory at any given moment, thus lightening the speaker's working memory load and decreasing the likelihood of interference between representations active in working memory. Informational cascades in language production allow for parallel processing at multiple levels of representation, and interactive feedback between levels of representation (e.g., phonological and lexical representations) can strengthen the activation of appropriate selections during lexical processing, thereby reducing competition from alternative representations and enabling faster lexical selection. Taken together, these features allow speakers to rapidly plan utterances and to do so during sentence articulation, enabling speakers to maintain a consistent, fluent rate of speech without long pauses between utterances that might present the opportunity for interruption.

In language comprehension, incremental, cascaded processing allows comprehenders to rapidly update hypotheses about the identity of the words appearing in the input, retrieve their meanings, and begin integrating those meanings into the representation of the message as a sentence unfolds (e.g., Allopenna et al., 1998; Yee & Sedivy, 2006), minimizing the need for long-term storage of the input in working memory prior to parsing. Furthermore, top-down

interactivity allows comprehenders to take advantage of contextual information to more efficiently process the bottom-up input, ruling out incongruent lexical candidates during word recognition (e.g., Dahan et al., 2000; Dahan & Tanenhaus, 2004; Magnusson et al., 2008) and pre-activating upcoming word representations even before they are encountered in the input (see Ito, 2024; Kuperberg & Jaeger, 2016; Ryskin & Nieuwland, 2023 for review). This dynamic interactive processing makes language comprehension more efficient and robust to noise in the input.

Given the importance of incremental and interactive cascades to contemporary psycholinguistic theories of language processing, understanding when and how these cascades develop is crucial for formulating theories of language acquisition and real-time language processing in children. Furthermore, understanding the flow of information in the developing language system can provide insight into the source of the adult system dynamics by revealing which features emerge early and thus may reflect the basic, fundamental properties of the language system as opposed to properties that emerge later as by-products of domain-general cognitive efficiency or extensive domain-specific experience.

1.1.3. Information flow in the developing language system

It is possible that the complex and dynamic informational cascades that underlie adult language processing are a fundamental consequence of the language system architecture that arises early in development. Indeed, incrementality, one of the key prerequisites for interactive cascades, emerges by the second year of life (Fernald et al., 2001; Swingley, 2009; Swingley et al., 1999). However, there are theoretical reasons to believe that the extent of cascading activation and interactivity within the language system may change over the course of

development. In particular, differences in language experience and domain-general limitations in children's cognition relative to adults may result in informational cascades (in top-down and/or bottom-up directions) that are limited, weaker, and/or more difficult to detect in young children. I discuss some of these differences and their potential effects on information processing below.

For instance, young children have more limited working memory capacities than adults (Chi, 1978; Cowan, 2017; Cowan et al., 2006; Dempster, 1981; Gathercole et al., 2004; Riggs et al., 2006; Schneider & Bjorklund, 1998; Simmering, 2012; inter alia). If working memory capacity constrains one's ability to simultaneously activate multiple representations at different processing levels, this could limit the extent of cascading and interactivity in the developing language system.

In addition, young children have weaker inhibitory abilities compared to adults and are more sensitive to interference in a variety of tasks (e.g., Bjorklund & Harnishfeger, 1990; Carter et al., 1995; Comalli et al., 1962; Dempster, 1992; Guttentag & Haith, 1978; Jerger et al., 2002; Jerger et al., 1988; Jerger et al., 1999; Ridderinkhof, 2002; Vurpillot & Ball, 1979). Such reduced inhibition may affect interactive informational cascades in two ways. First, it may weaken the influence of interactivity in young children by making it more difficult for them to inhibit or override representations activated at one level processing (e.g., representations activated by bottom-up information) in favor of influence from another level (e.g., top-down cues). For instance, in a study of homophone comprehension, Rabagliati et al. (2013) found that top-down cues from sentential context tend to be overridden by lexical associations in 4 year-old children. This difference could make it more difficult to detect evidence of interactive processing in young children.

Second, weaker inhibitory abilities could lead young children to be unable to effectively avoid or suppress ultimately unused representations that would become activated as a consequence of informational cascades and interactivity (e.g., the phonological forms of un-uttered lexical candidates or incorrect lexical predictions). Indeed, in language comprehension, children appear to consider competing lexical items for longer than adults do (e.g., Huang & Snedeker, 2011; McMurray et al., 2010; Sekerina & Brooks, 2007). Thus, if the language processing system is adaptively shaped by performance (evolutionarily or developmentally), informational flow between representations might be suppressed until inhibitory abilities are sufficiently mature. Similarly, such pressures could lead the interactive pathways for top-down and bottom-up influences in language comprehension and production, respectively, to mature more slowly than the bottom-up and top-down pathways minimally required for processing, reducing the level of interactivity in young children's language processing.

Moreover, young children have more limited language experience than adults. The representations within the language system and their relations to each other must be acquired on the basis of experience, and it has been suggested that the strength of these mappings and the level of interactivity within the lexical network increase with age (Bjorklund, 1995). Weaker mappings and reduced interactivity between representations early in life may make it more difficult for information to flow between representations both within and across levels of processing, limiting the extent of interactive cascades. It is possible that language processing only becomes efficient enough to allow for incremental interactive cascades once individuals have acquired sufficient language experience, which may occur later than early childhood.

Furthermore, in the realm of language comprehension, more limited language experience, immature executive functioning (Best & Miller, 2010; Mazuka et al., 2009), and/or slower

processing speed (Hale, 1990) may impede children's ability to construct and rapidly integrate higher-level representations into lower-level processing, thus reducing the amount of interactivity in the system. Indeed, there is evidence that young children (4–6 years) do not readily use top-down cues (e.g., plausibility, context) when resolving syntactic ambiguity (e.g., Kidd et al., 2011; Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008; Trueswell et al., 1999; Yacovone et al., 2021) and that school-aged children (approx. 7–12 years) do not use these cues to disambiguate homophones (Khanna & Boland, 2010; but cf. Hahn et al., 2015) or facilitate lexical access while reading (e.g., Joseph et al., 2008; Tiffin-Richards & Schroeder, 2020). In addition, there is evidence that rapid integration of top-down cues into language comprehension is more difficult for adults in a non-native (L2) language (e.g., Ito et al., 2018), which could predict comparable difficulty in young children, who also have more limited language experience. Nevertheless, there is evidence that in naturalistic tasks, five to ten-year-old children do use top-down cues to facilitate lexical processing (Levari & Snedeker, 2024), suggesting that the pathways required for interactive, cascading activation flow are already in place in this age range, even if they appear limited in their application or scope across tasks.

By exploring whether cascaded processing and interactivity are present in the developing language system by the age of four to five years, this dissertation can place constraints on our theories of when and how these capacities emerge, potentially shedding light on the innate properties of human language processing. I focus on the four to five-year-old age range, as children of this age are already proficient speakers of their native language but are in many ways still cognitively and linguistically immature. In particular, while four and five-year-old children have relatively large vocabularies (recognizing at least 10,000 words by the age of five; Schipley & McAfee, 2015) and are able to produce mostly well-formed sentences with high intelligibility,

their language knowledge and experience is still more limited than that of adults, and their language processing has been shown to be non-adult-like in several ways — for instance, displaying more speech errors and slower word production times (e.g., Cycowicz et al., 1997; D'Amico et al., 2001), more interference from competing lexical items (e.g., Huang & Snedeker, 2011; McMurray et al., 2010; Sekerina & Brooks, 2007), and more limited integration of top-down cues (e.g., e.g., Kidd et al., 2011; Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008; Trueswell et al., 1999; Yacovone et al., 2021). This age range thus serves as an important test case of the developing language system, allowing me to test whether the developing language system produces and comprehends language in the same way as the adult system.

1.1.4. Tracking moment-to-moment processing through the eyes: Moving the visual world into the virtual world

Visual world eye-tracking has emerged as an important tool for studying real-time language processing. In the visual-world paradigm, participants are presented with a visual display (typically a set of images), and their eye-movements are recorded as they listen to or produce an utterance. This paradigm leverages the temporal link between visual attention and language processing to assess how participants process linguistic information in real-time: Individuals systematically look to referents or associates of the words they hear (e.g., Cooper, 1974; Tanenhaus et al., 1995) or are planning to produce (e.g., Griffin & Bock, 2000; Meyer et al., 1998), and the timing of these saccades is tightly linked to the onset of the corresponding linguistic information (e.g., Cooper 1974; Griffin & Bock, 2000). Variations of this paradigm have been used to probe language comprehension and production at multiple levels of processing, including lexical processing (see Huettig et al., 2011 for review). In fact, visual

world eye-tracking research has been instrumental in establishing the presence of incremental cascades and interactivity during lexical processing in language comprehension (e.g., Allopenna et al., 1998; Dahan & Tanenhaus, 2004; Ito et al., 2024; Yee & Sedivy, 2006; among many others).

Visual-world eye-tracking is a particularly valuable tool for assessing these same features in the developing language system (in fact, I apply this approach in Chapter 3). This paradigm takes advantage of children's spontaneous responses to language input to reveal how the input is processed, without requiring lengthy set-up (cf. electroencephalography), meta-linguistic reasoning (cf. grammaticality judgments, lexical decision tasks), reading ability (cf. self-based reading), or complicated tasks (cf. picture–word interference). One limitation of eye-tracking studies is that they have traditionally required specialized eye-tracking devices that are housed within university labs (e.g., infrared eye-trackers; Tobii Pro, 2021; SR Research 2021). Infrared eye-trackers are highly accurate in both the temporal and spatial domains, providing coordinate-level estimations of participants' gaze locations (e.g., Ehinger et al., 2019; Nyström et al., 2021), however they are often relatively expensive and not easily portable. Limiting data collection to lab settings can make it more difficult to recruit large, diverse samples and to study languages not spoken near a researcher's home institution (see Henrich et al., 2010 for the importance of sample diversity).

Webcam eye-tracking techniques offer the potential to alleviate this limitation. These techniques use video recordings of participants to estimate their eye gaze location, either with automated gaze-estimation algorithms (e.g., Erel et al., 2022; Fraser et al., 2021; Papoutsaki et al., 2016; Valenti et al., 2009; Valliappan et al., 2020; Xu et al., 2015) or by hand-annotating gaze direction for each frame of the recorded videos (e.g., Ovans, 2022; Slim, Kandel et al.,

2024). These methods involve equipment that is easily portable (i.e., a laptop with a webcam) and that participants may have in their own homes. Consequently, webcam eye-tracking can allow researchers to recruit participants outside of lab settings, such as in museums, schools, or field sites (i.e., anywhere that they can bring a laptop) or over the internet. These approaches have benefits for child participants and their families, who are freed from needing to travel to the lab and (in the case of web-based experimentation) are able to complete the experiment in the comfort of their own homes, where children may feel more at ease than in unfamiliar lab settings.

Nevertheless, webcam eye-tracking techniques are less precise than the infrared eye-trackers typically used in the lab; automatic gaze-estimation algorithms have reduced spatiotemporal accuracy compared to infrared eye-trackers, and hand annotation does not provide coordinate-level gaze estimates (see Slim, Kandel et al., 2024 for discussion). Webcam eye-tracking may be even less accurate when estimating gaze for child participants, whose faces are smaller than those of adults and who are less likely to remain still in an optimal position throughout the duration of a task, making their eyes more difficult to see in the webcam view. In fact, even infrared eye-trackers are less accurate at estimating child gaze (Dalrymple et al., 2018). Moreover, children may be more prone to distraction when completing an experiment outside of a controlled lab setting. These factors may make webcam eye-tracking data too noisy or imprecise to carry out the types of experiments language researchers use to investigate features of moment-to-moment language processing.

Chapter 4 investigates the suitability of webcam eye-tracking approaches for conducting language research with school-aged children. As part of this investigation, I test whether these methods have the spatiotemporal sensitivity necessary to detect evidence of incrementality in

young children's lexical processing, providing insight into the methods' abilities to probe real-time processing effects.

1.1.5. Dissertation goals

The present dissertation examines the dynamics of information flow in the developing language system, using lexical processing as a case study. Across three papers, the dissertation asks whether the language system of four to five-year-old children behaves similarly to the adult system (showing rapid incremental, cascading, and interactive transfer of information across levels of representation), and it probes the limitations of webcam eye-tracking approaches to explore these questions. In particular, the dissertation focuses on how word representations are accessed and the ways that this access interacts with processing at other levels. This question is addressed from the perspectives of both language production (focusing on the temporal relationship between lexical selection and phonological encoding processes) and language comprehension (focusing on how top-down information influences the activation of lexical representations and guides bottom-up processing).

Paper 1 (Chapter 2) investigates whether five-year-old children show informational cascades between the processes of lexical selection and phonological encoding in language production, which would suggest that that the child language system involves the same rapid, cascading transfer of information across levels of presentation as the adult system. Paper 2 (Chapter 3) asks whether informational cascades in early childhood are interactive, moving in both top-down and bottom-up directions, similar to the adult system; the study uses visual-world eye-tracking to test whether four and five-year-old children are able to use information at higher levels of representation (e.g., sentence context) during language comprehension to pre-activate

upcoming lexical representations and to constrain processing of the bottom-up input, thereby providing evidence of interactivity in the developing language system. Paper 3 (Chapter 4) assesses whether eye-tracking paradigms of the type used in Paper 2 can be conducted in web-based settings using webcam eye-tracking techniques, which would allow language acquisition researchers to extend their reach when studying moment-to-moment processing in children; the study investigates whether webcam eye-tracking techniques are sensitive enough to detect evidence of incrementality during lexical processing in language comprehension.

Taken together, these papers inform how and when the complex dynamics of the language system develop as well as the methods language researchers can use to explore real-time processing in young children. These questions are important not only for language researchers who seek to understand what properties of the human language system are innate but also more generally for developmental psychologists who are interested in the mechanisms of informational flow through the cognitive system and how they develop. The results presented in this dissertation suggest that incremental interactive cascades arise early in language development, suggesting that they are fundamental properties of the language system and the mind more generally.

Chapter 2

[Paper 1]

Cascaded processing develops by five years of age:

Evidence from adult and child picture naming

Margaret Kandel & Jesse Snedeker

Published 2023 in Language, Cognition, and Neuroscience

[https://doi.org/10.1080/23273798.2023.2258536]

2.1. Introduction

Psycholinguistic theories break the process of producing a word into several steps (e.g., Butterworth, 1989; Dell, 1986; Friedmann et al., 2013; Garrett, 1980; Levelt, 1989; Levelt, 2001; Schriefers & Vigliocco, 2015; inter alia). Contemporary models have three major stages prior to articulation planning: (i) conceptual processing (determining the lexical concept to be conveyed; e.g., Levelt et al., 1999), (ii) lexical selection (identifying the lexical representation from the mental lexicon that will best convey the lexical concept), and (iii) phonological encoding (accessing the phonological form of the word to be uttered). A central question in language production research is how these processing levels interact and how information is passed between them.

There is compelling evidence that adult speakers carry out these processes in a cascaded fashion, with one process beginning before the earlier one is complete. Specifically, speakers can begin accessing the phonological forms of the lexical items under consideration even before a particular lexical item has been selected (e.g., Costa et al., 2000; Cutting & Ferreira, 1999; Jescheniak & Schriefers, 1998; Morsella & Miozzo, 2002; Peterson & Savoy, 1998; Rapp & Goldrick, 2000; Starreveld & La Heij, 1995). This process is frequently modelled as activation spreading between networks of interconnected nodes (e.g., Bloem & La Heij, 2003; Caramazza, 1997; Collins & Loftus, 1975; Dell, 1986; Dell et al., 1997; Levelt et al., 1999; Roelofs, 1992). Cascaded processing has a potential functional benefit: Speakers can get a head start on later levels of processing while finishing up the earlier ones, allowing for rapid and fluent production. In addition, cascaded processing allows representations at lower levels to become active while representations are still under consideration at a higher level, opening the door for processing at

lower levels to influence selection at the higher level through feedback loops (e.g., Dell, 1986; Dell et al., 1997; Dell & O'Seaghdha, 1991, 1992; Harley, 1993).

Given the centrality of cascaded processing to our theories of production, understanding the development of this ability is critical. Is cascading activation a fundamental property of the language production system, a consequence of its architecture that is present from early in development? Or does it emerge gradually with experience, appearing only after processing speed (or efficiency) approaches adult-like levels, thus freeing up resources for more processes to occur simultaneously? Curiously, despite the vast literature investigating activation flow in adult word planning, there is little work that explores how and when this ability develops (but see Jescheniak et al., 2006 discussed below). The present paper addresses these questions by comparing single word production in adults and five-year-old children using a picture naming paradigm.

2.1.1. Evidence for cascading activation in adult language production

Evidence for cascading activation in adult language production is primarily derived from two major sources: speech error analyses and reaction time studies.

Cascading activation models accurately capture the distributions of speech error types in adults with and without impaired lexical access in a variety of tasks including picture naming, word–picture mapping, and word repetition (e.g., Dell et al., 1997; Foygel & Dell, 2000; Rapp & Goldrick, 2000; Schwartz et al., 2006). Critically, this theory correctly predicts the frequency of *mixed errors*, substitution errors that are both semantically and phonologically related to the intended word (e.g., saying *rat* in place of *cat*). In both spontaneous and experimentally-elicited speech, adults produce mixed errors at a higher rate than would be expected given the base rates

of purely semantic and purely phonological errors (e.g., Blanken, 1998; Butterworth, 1981; Dell & Reich, 1981; Harley, 1984; Martin et al., 1989; Martin et al., 1996; Rapp & Goldrick, 2000; Stemberger, 1983). The high likelihood of mixed errors arises as a natural consequence of cascading activation (e.g., Rapp & Goldrick, 2000) but is not predicted by discrete, serial word planning models that assume that a word's phonological form is only activated once its lexical representation has been selected for articulation (e.g., Butterworth, 1992; Garrett, 1980; Levelt et al., 1999; Roelofs, 1992; Schriefers et al., 1990.[1]

Additional evidence for cascading activation comes from reaction time studies that explore when lexical and phonological representations become active during word planning. These studies often use the picture-word interference (PWI) paradigm, in which participants must name a picture accompanied by a distractor word (presented either visually or auditorily). Distractors that are semantically related to the picture referent generally result in slower naming times (e.g., Damian & Bowers, 2003; Damian & Martin, 1999; La Heij et al., 1990; Roelofs, 1992; Schriefers et al., 1990; Vigliocco et al., 2004; but cf. Mahon et al., 2007), an effect which is believed to reflect increased difficulty choosing the intended word during lexical selection due to competition from the semantic distractor (Levelt et al., 1999; Roelofs, 1992; but cf. Mahon et al., 2007 for an alternative proposal). When a distractor is phonologically related to the picture name, on the other hand, naming RTs tend to be shorter (e.g., Levelt et al., 1991; Meyer & Schriefers, 1991; Schriefers et al., 1990), which is thought to reflect activation of the phonemes in the target name shared by the distractor word (Damian & Martin, 1999; Dell & O'Seaghdha, 1992; Roelofs et al., 1996). There is evidence that phonological facilitation can occur within the

[1] Reconciling the mixed error effect with a serial model of lexical planning (e.g., Levelt et al., 1991) requires the assumption of a post-encoding editor (Baars et al., 1975; Butterworth, 1981; Kempen & Huijbers, 1983; Levelt, 1989).

same early time windows as semantic interference (e.g., Cutting & Ferreira, 1999; Damian & Martin, 1999; Jescheniak & Schriefers, 2001; Starreveld, 2000), suggesting that both lexical and phonological representations can be active at the same time, as predicted by cascading models. In addition, when distractors are both semantically and phonologically related to the target, there is an interaction between the two effects, a pattern which suggests that lexical selection and phonological encoding are not serial and independent (Damian & Martin, 1999; Starreveld & La Heij, 1995). Not all studies have found evidence of early phonological effects, however, (e.g., Schriefers et al., 1990), and these effects do not provide fully conclusive evidence of cascaded processing, as phonological effects in PWI paradigms are open to a range of explanations, including some in which the effect does not arise directly from production processes (Starreveld, 2000). Furthermore, it is possible to account for the interactions observed in PWI tasks between semantic and phonological effects within a serial framework of word planning (e.g., Roelofs et al., 1996).

More direct evidence for cascading activation in adult production is derived from RT studies that probe the phonological activation of un-uttered words, typically unselected alternative lexical candidates or words that are semantically related to the produced word. Evidence for the activation of words related to the produced word (semantically-mediated phonological activation) has been found in a number of tasks including priming paradigms (e.g., Peterson & Savoy, 1998), PWI (e.g., Abdel Rahman & Melinger, 2008; Jescheniak et al., 2005; Jescheniak et al., 2006; Jescheniak & Schriefers, 1998; Melinger & Abdel Rahman, 2013; Zhang et al., 2018; inter alia), and EEG paradigms (Jescheniak et al., 2003). Evidence has also been found for the activation of words related to a homophone of the produced word (Cutting & Ferreira, 1999). Additional evidence for phonological activation of un-uttered words comes from

bilingual picture naming (RTs are faster when the names in both languages are phonologically similar, suggesting both phonological forms are activated; Costa et al., 2000) and from picture–picture interference paradigms in which participants have to name one of two superimposed images (RTs are faster when the name of the un-named image is phonologically related to the produced name; e.g., Humphreys et al., 2010; Kuipers & La Heij, 2009; Mädebach et al., 2011; Meyer & Damian, 2007; Morsella & Miozzo, 2002; Navarette et al., 2017; Navarrete & Costa, 2009; Roelofs, 2008).

The present study uses a naming paradigm to explore an additional, unstudied, prediction of the cascading architecture: that phonological activation can begin while lexical selection is still underway, thereby resulting in interactions between the variables that influence each of the two processes (more below). The picture naming task is considerably simpler than the priming and interference tasks discussed above, thus we can investigate this prediction not only in mature speakers but also in young children, for whom the dynamics of lexical access have not been as extensively studied.

2.1.2. The limited evidence for cascading activation in child language production

Speech error and aphasia research suggests that the developing language production system is similarly organized to that of adults, with distinct stages for lexical selection and phonological processing (Friedmann et al., 2013). There is compelling (although limited) evidence for cascading activation in word planning in children over the age of seven. Jescheniak et al. (2006) observed evidence of semantically-mediated phonological activation in the picture naming behavior of second graders (aged 7;3–8;6). In an auditory PWI task, Jescheniak et al. (2006) found that second graders were slower to produce target picture names (e.g., *Mantel*

[coat]) in the presence of a distractor (e.g., *Honig* [honey]) that had the same phonological onset as a word that was semantically related to the target (e.g., *Hose* [trousers]) but was not phonologically or semantically related to the target itself or semantically related to the target-related word. Such an interference effect would be expected if activation cascades to the phonological forms of semantic associates of the target word (Jescheniak et al., 2005; Jescheniak & Schriefers, 1998). Jescheniak et al. (2006) did not find a similar effect in fourth graders (aged 9;4–10;8) or adults, leading them to propose that cascading activation is present across development but is easier to detect when the lexical planning process is more stretched out in time, as it is for the seven and eight-year-old children.

Other studies with school-aged children are suggestive of cascaded processing, even if they are open to alternative explanations. For example, children eight to eleven years old, like adults, show phonological facilitation effects in the same early time windows as semantic interference effects (e.g., Sieger-Gardner & Schwartz, 2008; Sylvia, 2017). In addition, like adults, school-aged children (eight to eleven years) are influenced by the phonological forms of context picture names in picture–picture interference paradigms (Sylvia, 2017), though Sylvia (2017) observed an interference effect from the distractor image rather than the facilitation effect that is commonly observed in adults (e.g., Morsella & Miozzo, 2002). Finally, Poarch and van Hell (2012) found that multilingual children aged five to eight years old (M age = 7.28 years, SD = 0.76) demonstrate a bidirectional cognate phonological facilitation effect between German and English, with faster RTs when the names in both languages are phonologically similar, which could indicate that both phonological forms become activated (or it could reflect a phonological frequency confound; see Costa et al., 2000 for discussion).

In contrast, the evidence of cascading activation for children under the age of seven is limited and weak. Three lines of research are potentially relevant. First, children between five and seven years of age show early phonological facilitation effects in an auditory version of the PWI paradigm (Jerger et al., 2002). These effects, however, are open to the same alternative explanations as the parallel effects in adults (e.g., Starreveld, 2000). Second, like adults, young children (1–5 years) produce mixed lexical substitution errors (Jaeger, 2005). Unfortunately, these analyses do not assess whether the frequency of substitution errors both semantically and phonologically related to the target is greater than what would be expected if phonological and semantic errors are independent (as they would be in a model with no informational cascade). Third, adult speech error models that include cascading activation (e.g., Foygel & Dell, 2000) can simulate error distributions in children five to eleven years old with quantitative shifts in model parameters across ages (Budd et al., 2011), suggesting that broadly similar processes are at work in adults and children. Evidence consistent with one model, however, does not rule out others. It is unclear whether the cascading model provides a better fit of the child error distribution than models with no informational cascade.

In sum, while a cascading architecture is likely present by around seven years of age, the evidence for cascading activation in word planning before this age is weaker and open to multiple interpretations. The present study explores whether cascading activation is present for five-year-olds, children who are proficient speakers of their native language but have had little formal schooling and are largely pre-literate.

2.1.3. Reasons why activation flow may differ in child language production

The cognitive and linguistic abilities of children five years of age and younger differ from older children and adults in several ways that could have implications for the development of cascaded processing.

Some of these differences might lead us to expect cascading activation to be limited, weaker, or more difficult to detect in children this young. For example, although five-year-olds are proficient speakers, they have considerably less experience with language than adults, in the sense that they have had many fewer years of using language and have, for example, smaller vocabularies (see Goulden et al., 1990; Nagy et al., 1985; Zareva et al., 2005 for estimates of the adult active working lexicon; see Shipley & McAfee, 2015 for child vocabulary estimates). Informational cascades arise when there is rapid processing that results in quick information transfer across levels of representation. In the case of lexical processing, these interacting representations (word forms and meanings) must be acquired on the basis of experience. Thus, it seems plausible that considerable experience might be required before processing becomes efficient enough to support this level of incrementality. Indeed, it has been suggested that both the strength of the mappings between nodes and the interactivity of the lexical network increases with age (Bjorklund, 1995).

Domain-general limitations in children's cognition could also limit or prevent cascaded processing. For instance, five-year-olds have more limited working memory than adults (Chi, 1978; Cowan, 2017; Cowan et al., 2006; Dempster, 1981; Gathercole et al., 2004; Riggs et al., 2006; Schneider & Bjorklund, 1998; Simmering, 2012; inter alia). If working memory places limits on children's ability to simultaneously activate multiple lexical representations and multiple phonological forms, this could result in less cascading activation in younger children.

Young children also have much weaker inhibitory abilities than adults: For example, they tend to be more susceptible to interference in Stroop tasks and other tasks requiring the suppression of a distractor (Bjorklund & Harnishfeger, 1990; Carter et al., 1995; Comalli et al., 1962; Dempster, 1992; Guttentag & Haith, 1978; Jerger et al., 2002; Jerger et al., 1988; Jerger et al., 1999; Ridderinkhof, 2002; Vurpillot & Ball, 1979). If multiple lexical and phonological representations are active at one time (as would occur within a cascading architecture), and children are unable to effectively inhibit candidates as new information comes in, then the costs of allowing activation to cascade across levels might be greater than the benefits. If this is the case, and if the processing system is adaptive (shaped by performance, over evolutionary or developmental time), then informational cascades might be suppressed until inhibitory abilities are more mature.

On the other hand, there are also reasons to think that one might find stronger or more widespread cascading activation in five-year-old children. First, if cascading activation is a basic property of cognitive architecture that cannot be suppressed (even when it is counterproductive) and if children have reduced inhibitory ability (see above), then the cognitive signatures of interactivity may be easier to detect in this population. A young child's inability to inhibit alternative lexical candidates might lead these candidates to be active for longer, allowing more time for information to cascade to the phonological level, providing a stronger signal of the cascade in measurements such as reaction times and speech errors. In contrast, older children and adults may more efficiently suppress the activation of non-target forms. This hypothesis could explain why Jescheniak and colleagues (2006) found evidence of semantically-mediated phonological activation in children seven to eight years old but not in older children or adults (Jescheniak et al., 2003; Jescheniak et al., 2006).

Second, young children's slower responses could result in more time for information to cascade prior to production. Five-year-old children typically take about 200 ms longer to name a picture than adults do (e.g., D'Amico et al., 2001 for five to six-year-olds; Jerger et al., 2002 for five to seven-year-olds). If this slowdown results from longer lexical selection times (as children consider competing lexical candidates), this additional time could allow more activation to cascade from the active lexical representations and their phonological forms, strengthening the activation of these forms beyond the levels they would typically reach in adults (assuming that the processes leading to the spread are not also slowed down). The effect of this slowdown thus might be similar to the effect of reduced inhibition described above. This increased activation of phonological forms could allow for more easy detection of cascaded activation, such as through semantically-mediated phonological effects.

By looking for signatures of cascading activation in five-year-old children, we can begin to target the questions of when and how this capacity emerges.

2.1.4. The present study

In the present study, we perform a side-by-side comparison of the mature and developing language production systems, comparing word planning by adults and five-year-old children in a picture naming task. We investigate the effects of image codability (name agreement) and name frequency on picture naming response time. These factors are commonly manipulated in language production experiments (e.g., Ferreira & Pashler, 2002; Griffin, 2001; Lee et al., 2013; Momma, 2021; inter alia), elicit robust effects across multiple languages and age groups (e.g., Bates et al., 2003; D'Amico et al., 2001; Johnson, 1992), and their effects are believed to index psychological processes within individuals during word planning. Crucially, they have been

argued to influence the processes of lexical selection and phonological encoding, respectively, thereby allowing us to tap into the interplay between these processes in our two populations.

The codability of a referent serves as a measure of name agreement (e.g., Snodgrass & Vanderwart, 1980). Highly codable referents have a high level of name agreement (i.e., fewer alternative names applied to them), whereas referents with low codability have less name agreement (i.e., more possible names that can describe them). Speakers are faster to name pictures with higher codability than those with lower codability. Codability effects have been observed in both adults (e.g., Lachman, 1973; Lachman et al., 1974; Lachman & Lachman, 1980; Paivio et al., 1989) and in children, as young as four years of age (e.g., Butterfield & Butterfield, 1977; Johnson, 1992; Johnson & Clark, 1988). The codability effect is generally thought to influence the process of lexical selection, rather than visual processing or picture identification (Alario et al., 2004; Griffin, 2001; Johnson, 1992; inter alia). When there are multiple possible names for a speaker to choose from, these alternatives are simultaneously activated and compete with each other for selection. The greater the number of alternatives, the more competition, and thus the longer it takes for the speaker to resolve this competition and select a label. Recent research by Balatsou et al. (2022) supports this interpretation, finding evidence that name agreement is predictive of lexical co-activation within-speakers (though population-level measures may overestimate within-speaker variability).[2]

Word frequency has also been found to influence picture naming time in both adults (e.g., Bates et al., 2003; D'Amico et al., 2001; Jescheniak & Levelt, 1994; Lachman, 1973; Lachman et al., 1974; Oldfield & Wingfield, 1965; inter alia) and children (e.g., D'Amico et al., 2001 for

[2] It is important to note, however, that codability effects, while commonly attributed to co-activation at the lexical level, may not exclusively reflect an influence on lexical decision; name agreement may also influence processes prior to lexical decision such as conceptual access.

five to six-year-olds). The frequency effect in word production is generally viewed as operating on the level of phonological encoding (e.g., Jescheniak & Levelt, 1994; see Griffin & Bock, 1998 for an overview of evidence in favor of this interpretation from RT and speech error studies; see Bates et al., 2003 for additional discussion of the loci of frequency RT effects), with more frequent word forms having higher resting activation, thereby allowing them to be accessed and selected more quickly than low frequency phonological forms during encoding. There is, however, some evidence that frequency may influence the lexical selection process in addition to affecting the selection of phonological form (e.g., Finocchiaro & Caramazza, 2006; Jescheniak & Levelt, 1994; Johnson et al., 1996; Kittredge et al., 2008; Strijkers et al., 2010), highlighting the complexity of attributing variables to specific processes within language production. Nevertheless, the effect of word form frequency is independent from that of conceptual frequency (Bates et al., 2003), and the effect of frequency on phonological encoding appears to be stronger and more reliable than its effect on lexical selection (Griffin & Bock, 1998).

Given that the codability effect has its locus in lexical selection and the frequency effect largely impacts phonological encoding, investigating how these two factors influence picture naming times allows us to investigate how these processes relate to one another during production. A statistical interaction between these effects could suggest that these processes interact, as is predicted by a cascading model of lexical planning. In contrast, factors that influence discrete, serial processing stages predict additive effects (Sternberg, 1969, 2001; but cf. Stafford & Gurney, 2011; Thomas, 2006). To our knowledge, such an interaction has not been previously reported for either child or adult picture naming. Critically, this is not because researchers have looked for interaction and failed to find it. The few picture naming studies we have found that directly explore both factors simply do not report on the presence or absence of

an interaction in codability and frequency's influence on naming timing (e.g., Alario et al., 2004; Bates et al., 2003; Cycowicz et al., 1997; D'Amico et al., 2001; Lachman et al., 1974).[3]

Study 1 provides a side-by-side comparison of codability and frequency picture naming effects in five-year-olds and adults, assessing the similarity between adult and child lexical selection and phonological encoding processes. We replicate previously-observed RT effects of codability and frequency in our two populations. Critically, we also find an under-additive interaction between these effects. This pattern could suggest that phonological encoding begins before lexical selection ends, reducing the cost when both processes are slow — exactly what we would expect if there is an incremental cascade of information across these levels of representation. To further explore the effects of codability and frequency in our two populations, we fit ex-Gaussian distributions to our data to investigate how our manipulations influence the RT distributions (e.g., Staub, 2010), testing whether codability and frequency have qualitatively similar or different RT effects from each other and whether they influence adult and child participants similarly, suggesting comparable underlying processes. Study 2 tests the reliability of the interaction effect observed in Study 1, looking for comparable effects across languages in the adult naming data collected by Bates et al. (2003). Study 3 investigates how such an interaction might arise by simulating how the relationship between lexical selection and phonological encoding can influence RT in both serial and cascading activation architectures.

[3] One exception we have found is an adult sentence production study by Spieler and Griffin (2006). Their experiment elicited sentences in the form *The A and the B is above the C*. The researchers manipulated the frequency (high, low) and codability (high, medium) of critical items that appeared in either the B or C position (the item in A always had high codability). They observed an interaction between the frequency and codability of the critical items on the latency between the onset of A and the onset of the critical item. This interaction is not in the direction we observe, however: they observed an over-additive effect of frequency for medium codable items compared to highly codable items (latencies were especially slow for low frequency, medium codable items). While not extensively discussed in Spieler and Griffin (2006), they similarly attribute such an effect to cascading activation (see discussion in Griffin & Bock, 1998 about how increased constraint in word choice may attenuate the influence of frequency).

2.1.5. Data availability

The data, analysis code, and Supplementary Materials for all studies in this paper are available from: https://osf.io/myrtg/.

2.2. Study 1: A picture naming experiment with adults and five-year-old children

Study 1 investigated image codability and name frequency effects on picture naming RT in adults and five-year-old children. There were two sets of questions we sought to answer in this study:

- Q1 (*RT effects of codability and frequency*): Are five-year-olds, like adults, slower to name pictures with low name agreement and low name frequency? Do the effects of codability and frequency interact in both populations?
- Q2 (*Influences of codability and frequency on the RT distribution*): Do the codability and frequency manipulations influence RT distributions similarly? Are these RT distribution patterns qualitatively similar for adult and child responses (suggesting similar underlying processes) or different (suggesting changes in the lexicon over the course of development)?

2.2.1. Methods

The experiments in this study were approved by the Harvard University-Area Committee on the Use of Human Subjects (Protocol # 12718). The experiment methods were preregistered on OSF for both our adult (https://osf.io/hwtzs) and child (https://osf.io/3zcp8/) participants. The

categorical analysis of codability and frequency in Q1 was preregistered. The remaining analyses in this paper were exploratory.

2.2.1.1. Participants

The participants were 48 adults (M age = 19.9 years, SD = 1.3; range = 18–23 years; 38 F, 10 M) and 25 children (M age = 5.47 years, SD = 0.29; range = 5;0–5;11; 7 F, 18 M). The number of child participants was determined via power analysis based on the effect sizes in the adult categorical analysis.[4] All participants were native, monolingual American English speakers. Adults were recruited from undergraduate classes at Harvard University and received partial course credit for their participation. Adult participants provided informed written consent to participate in the study. Children were recruited from the Harvard Laboratory for Developmental Studies database; they were given a small toy for participating, and their parents were given a $5.00 travel reimbursement. Informed written consent was received from the parent or guardian of the child participants for the child's participation. Six additional adults were tested but excluded from the analysis due to technical errors (1) or because they were early bilinguals (5). Three additional children were tested but excluded due to trial loss of over 50% (2) or because they were bilingual (1).

2.2.1.2. Materials

Participants viewed stimulus images from the BCBL MultiPic databank (Duñabeitia et al., 2018) and the Snodgrass and Vanderwart "Like" Objects (Rossion & Pourtois, 2004). The images were colorized digital images with black outlines. Adults saw and named 200 pictures.

[4] The power analysis was performed using the random effects structure specified in our preregistration.

For the child experiment, we reduced the number of pictures to 120 so that the children would be more likely to complete the experiment. When selecting the images for the child experiment, we excluded items that received a large number of non-synonymous name responses in the adult data set, suggesting that the image was difficult to identify. For ease of comparison, the present analyses include only the responses to the 120 images named by both children and adults. We identify places where the result patterns differed in the full set of adult responses.

The experiment had a 2 × 2 within-subjects manipulation of image codability (high, low) and image name frequency (high, low), resulting in four conditions: *High Codability, High Frequency* (e.g., apple), *High Codability, Low Frequency* (e.g., cactus), *Low Codability, High Frequency* (e.g., sofa/couch), and *Low Codability, Low Frequency* (e.g., spaceship/UFO). To ensure that we could get responses that varied along the codability and frequency dimensions, for the adult stimulus set, we selected 50 images that we expected to fall into each of the four quadrants of our design (see Supplementary Materials for details). None of the selected items were intended to elicit names that were conceptually or grammatically plural. After collecting the adult data (but prior to analysis), images were assigned to codability and frequency categories (see *§2.2.2.1. Q1 data analysis procedure*). The 120 images named by both adult and child participants included 30 *High Codability, High Frequency* items, 29 *High Codability, Low Frequency* items, 29 *Low Codability, High Frequency* items, and 32 *Low Codability, Low Frequency* items.

2.2.1.3. Procedure

Each participant was tested individually in a single 20–30 min session. The stimulus images were shown on a Tobii T-60 remote eye-tracker. The session began with four practice trials (always in the same order) followed by the experimental trials, which were randomized.

At the beginning of each trial, a fixation cross appeared in the center of the screen for 500 ms and was then replaced by a single stimulus image against a white background. For adult participants, the stimulus image remained on screen for four seconds before the trial ended. This procedure was slightly modified for the child participants. Given that similarly-aged children have been found to produce picture names more slowly than adults in comparable paradigms (e.g., D'Amico et al., 2001 for five and six year-olds), the maximum response time was increased to five seconds. As children are prone to loss of attention and/or fatigue in longer experiments, in order to minimize pauses and reduce the overall experiment duration, the experimenter advanced to the next trial after the participant named the object or indicated that they did not know the name.

Participants were instructed to name each image as quickly and as accurately as possible using a single word, to speak clearly, and to avoid producing any other words (e.g., articles) or sounds (e.g., clearing of the throat or filler sounds such as "umm") before giving a name. Audio responses were recorded through a microphone headset worn by the participant. The recording for each trial started when the stimulus image appeared on the screen and stopped when the image disappeared. Onset latencies were determined from these recordings.

2.2.1.4. Data exclusion & calculating onset

Responses were excluded from the analysis if the participant did not speak or did not name the object, if the audio recording did not contain the complete response, if the response contained more than a single name (renaming), if the response contained a false-start or repeated the start of the name, if the response contained a prenominal verbalization (e.g., an article, "that's a … ", etc.),[5] if the response contained a prenominal sound (e.g., clearing of the throat, cough, "ummm", speech from the experimenter, participant, or parent) preventing the determination of the name onset time by the forced-aligner (see below), or if the onset time was otherwise incalculable from the recording (e.g., due to poor audio quality or background noise). We added an additional exclusion criterion beyond those specified in our preregistration to omit responses with post-nominal descriptions (e.g., "apple with a leaf", "baby touching its toes"), though we allowed responses in the form N of N ("ball of yarn", "jar of honey", "spool of thread", "loaf of bread"). Multi-word responses were excluded from the continuous analyses, as they lacked SUBTLEX-US frequency measures (Brysbaert & New, 2009) (see *§ 2.2.2.1. Q1 data analysis procedure*).

For each response, the name was transcribed and speech onset time (measured from image onset) was determined using the Montreal Forced Aligner v1.0.0 (McAuliffe et al., 2017). Responses with prenominal sounds were flagged during the transcription process, and their alignments were checked; if onset time was identified by the forced aligner as the onset of the prenominal sound rather than the onset of the name, the response was omitted from analysis.

[5] The adult preregistration lists the use of an adjective as a potential example of prenominal verbalization. Given that some of the stimulus images elicited responses with adjectives that modified the head noun such that the response could potentially be considered a single lexical item (e.g., a compound noun) with a different meaning than the head noun (e.g., school bus, steering wheel, candy bar), we decided to allow adjectives in the responses.

2.2.2. Q1: RT effects of codability and frequency

The first question we sought to answer in our experiment was whether codability and frequency influence naming RT in both the adult and child responses. We looked for previously-observed slowdowns in cases of low codability and low frequency, and we also looked for an interaction between the two effects.

2.2.2.1. Q1 data analysis procedure

All statistical analyses reported in the present paper were performed in R v 4.1.0 (R Core Team, 2021). Adult and child data were analyzed separately.

2.2.2.1.1. Categorical analysis.

Items were categorized into codability and frequency categories based on their name agreement and the frequency of the dominant names applied to them (their "target names") in the adult responses. When computing item name agreement (codability) and target names, we included all responses that gave a single complete label to the image, even if that response was not ultimately included in the analysis (e.g., responses including a determiner before the name; see *§2.2.1.4. Data exclusion and calculating onset*).

An item's codability category (high, low) was determined based on its H score in the adult data. The H statistic is a measurement of name agreement calculated using the formula $H = \sum_{i-1}^{k} p_i log_2(1/p_i)$, where k is the number of different names given to the image, and p_i is the proportion of participants providing a specific name (e.g., Snodgrass & Vanderwart, 1980). The lowest possible value of H is 0, indicating perfect name agreement. The maximum H is

achieved when each participant gives a different response and varies based on the number of participants in the experiment. Items were categorized based on the ratio between each image's H score and the maximum possible score in the experiment. The proportion of participants giving each name in the H score calculation was determined based on the number of responses included in the computation (rather than the total number of participants in the experiment). Items with H score ratios less than or equal to 0.06 were assigned to the high codability category, and items with H score ratios greater than 0.06 were assigned to the low codability category. The high codability items (n = 59) had an average H score of 0.07 (SD = 0.11) (M H score ratio = 0.01, SD = 0.02), and the low codability items (n = 61) had an average H score of 1.39 (SD = 0.57) (M H score ratio = 0.25, SD = 0.10).

An item's frequency category (high, low) was determined based on the frequency of its target name. Adult target names were used as the item target names for both the adult and child analyses. A frequency score was calculated for each item as a natural log transformation of its raw frequency in the SUBTLEX-US corpus (Brysbaert & New, 2009) [ln(1 + *raw frequency*), where *raw frequency* = words per million]. Items whose target names had frequency scores greater than 3.00 were assigned to the high frequency category, and items whose target names had frequency scores less than or equal to 3.00 were assigned to the low frequency category. The high frequency items (n = 59) had an average frequency score of 4.14 (SD = 0.90), and the low frequency items (n = 61) had an average frequency score of 1.74 (SD = 0.63).

We categorized the items based on adult measures because these measures are based on a larger number of responses (and are thus potentially more stable than corresponding child measures), and using the same categories for both populations allows for a direct comparison of their naming RT in response to the same items in the categorical analysis. Furthermore, the adult

codability and frequency measures used for item categorization are significantly correlated with the corresponding measures from the child data. Specifically, the H scores from the adult experiment ("adult H scores") and the H scores from the child experiment ("child H scores") had a Pearson's r of 0.67 ($t(118) = 9.79$, $p < 0.0001$). Similarly, there was a robust correlation between the items' frequency scores obtained from SUBTLEX-US ("frequency scores") and the frequency of their target names in the utterances of children aged 36 months and older the CHILDES corpus (MacWhinney, 2000) ("child frequency scores") (Pearson's $r = 0.81$, $t(118) = 14.96$, $p < 0.0001$).

The items in the high and low codability and frequency categories varied only along the dimensions we intended to manipulate. We confirmed this by conducting Type II Analyses of Variances (ANOVA) (performed using the R package {car} v3.0-11; Fox & Weisberg, 2019) and post-hoc pairwise-comparison tests (performed using {emmeans} v1.6.2-1; Lenth, 2019). The high and low adult codability items differed significantly based on adult H score ($t(118) = -17.42$, $p < 0.0001$) but did not differ significantly based on target name frequency score ($t(118) = 0.45$, $p = 0.66$). The high and low frequency items differed significantly based on target name frequency score ($t(118) = 17.06$, $p < 0.0001$) but not on adult H score ($t(118) = -1.21$, $p = 0.23$).

Table 2.1 shows the codability and frequency measures of our items broken down by the four experiment conditions (for additional properties of the items and their correlations with these measures, see Supplementary Materials). Critically, we succeeded in orthogonally manipulating codability and frequency across the four cells of our 2 × 2 design (confirmed via post-hoc pairwise-comparison tests with Tukey p-value adjustment; a full table of the pairwise comparison results is available in the Supplementary Materials). All of our codability and frequency measures reliably vary by condition ($F(3) \geq 35.80$, p's < 0.0001). We found

significant differences in adult H score only between conditions with different codability categories ($|t(116)| \geq 10.95$, p's < 0.0001), and we found significant differences in adult frequency score only between conditions with different frequency categories ($|t(116)| \geq 10.56$, p's < 0.0001).

The pattern of values for the child measures was more complicated. The child measurements differed significantly and substantially along the dimension we sought to manipulate: Child H score differed between conditions with different codability categories ($|t(116)| \geq 2.66$, p's < 0.05), and child frequency scores differed between conditions with different frequency categories ($|t(116)| \geq 9.43$, p's < 0.0001). We did find, however, that the child H score also differed between the *High Codability, High Frequency* and *High Codability, Low Frequency* conditions, with higher child H scores (less name agreement) when frequency was low ($t(116) = -5.28$, $p < 0.0001$). In addition, the child frequency scores differed between the *High Codability, High Frequency* and *Low Codability, High Frequency* conditions, with higher child frequency scores in the high codability group ($t(116) = 2.75$, $p = 0.03$). These imbalances follow the direction of a previously-observed relationship between codability and frequency: Name agreement tends to be higher for items with high frequency names (e.g., Bates et al., 2003). This pattern of differences could potentially result in an over-additive interaction in the child data analysis, as the *High Codability, High Frequency* is higher on the child measures of both of our manipulated dimensions than the other conditions. We address this concern in the continuous analysis.

Table 2.1: *Properties of the codability × frequency conditions.* Mean adult H score, child H score, adult frequency score, and child frequency score for item target names in the four codability × frequency conditions. SD in parentheses.

Condition	Adult H Score	Child H Score	Frequency Score	Child Frequency Score
High Codability, High Frequency	0.02 (0.06)	0.20 (0.31)	4.34 (0.95)	6.17 (1.27)
High Codability, Low Frequency	0.11 (0.13)	1.07 (0.81)	1.58 (0.57)	2.39 (0.68)
Low Codability, High Frequency	1.30 (0.59)	1.51 (0.61)	3.94 (0.80)	5.40 (0.89)
Low Codability, Low Frequency	1.47 (0.56)	1.76 (0.69)	1.89 (0.64)	2.82 (1.28)

We performed linear mixed effects analyses on log-transformed onset time (in log milliseconds) using the {lmerTest} package v3.1-3 (Kuznetsova et al., 2017). The linear mixed effects models used in the categorical analyses contained fixed effects for codability category (high, low), frequency category (high, low), trial number, and target name syllable count, with an interaction between frequency category and codability category and random slopes for codability category, frequency category, and their interaction by participant as well as a random intercept

by item.[6] The fixed effects of trial number and target name syllable count were intended to control for effects of fatigue (e.g., D'Amico et al., 2001) and word length (e.g., D'Amico et al., 2001; Johnson et al., 1996; Székely et al., 2003; Székely et al., 2005; see Bates et al., 2003 for cross-linguistic differences in length effects) that increase picture naming RT. As the fixed effects of codability category and frequency category were entered into our analysis models with an interaction, to estimate the overall influence of these variables on RT, we used effects coding for these variables in the regression (e.g., Hardy, 1993), allowing us to compare the influence of the variables on the grand mean RT (analogous to main effects). To compute pairwise comparisons across levels of the categorical variables, we repeated the analyses with the same model structures using dummy-coding of the codability and frequency variables.

2.2.2.1.2. Continuous analysis.

We addressed questions that arose from the results of the categorical analysis in an exploratory analysis (the "continuous analysis") that we had not pre-registered. To minimize the likelihood of false positives, wherever possible, we constrained the continuous analysis based on the decisions we made in the categorical analysis.

In this analysis, we investigated whether the data patterns observed in the categorical analysis persisted when we used continuous measures of codability and frequency. This analysis was intended to allay concerns that data patterns observed in the categorical analyses may be due to any departures from a perfectly orthogonal manipulation (particularly for the child data set). Paralleling the reported categorical results, we computed linear mixed effects models using the

[6] We altered the model structure specified in the preregistration to (i) make our model estimates more conservative and (ii) aid in model convergence. We added a random intercept for item as well as an interaction term in the random slope by subject, and we omitted the random effect of trial (an investigation of participant slopes for trial suggested minimal variation in trial slope by participant).

{lmerTest} package v3.1-3 (Kuznetsova et al., 2017) to analyze log-transformed onset time (in log milliseconds). When possible, the linear mixed effects models for the continuous analysis had the same effects structure as the categorical analysis, except that codability category and frequency category were replaced with H score and frequency score (any deviations from this structure are noted in the results). For maximum precision, we used the syllable counts and frequency scores of the individual responses rather than the items' target names. The adult analysis used adult H scores. We analyzed the child data using both child and adult H scores; although the adult H scores reflect the name variation in the language input that children hear and were derived from a greater number of responses (as mentioned above), the child H scores may be a more direct reflection of child naming behavior. Given the correlation between the frequency scores from SUBTLEX-US and from CHILDES (see above), we used the SUBTLEX-US frequency scores for both analyses, as they are based on a larger corpus of utterances.

2.2.2.2. Q1 results

2.2.2.2.1. Participant responses

For the 120 stimulus items that were viewed by both adult and child participants, we recorded 5760 adult responses and 3000 child responses. 74 adult responses and 593 child responses were excluded from the analyses based on the criteria above (see *§2.2.1.4. Data exclusion & calculating onset*). The distributions of false start, repeated start, renaming, and no response errors by codability and frequency category in the adult and child data are available in the Supplementary Materials. An additional 66 adult responses and 70 child responses were

omitted from the continuous analyses because frequency counts were not available for them in SUBLTEX-US.

Replicating the results of previous studies (e.g., D'Amico et al., 2001), our child participants were in general slower to name the stimulus images than our adult participants: The average adult RT was 1034 ms (SD = 471 ms), and the average child RT was 1319 ms (SD = 632 ms). The distribution of onset times by condition are given in Figure 2.1 (adult data) and Figure 2.2 (child data).

Figure 2.1: *Adult RTs by condition.* Each point represents a response. The black diamonds indicate mean RT.

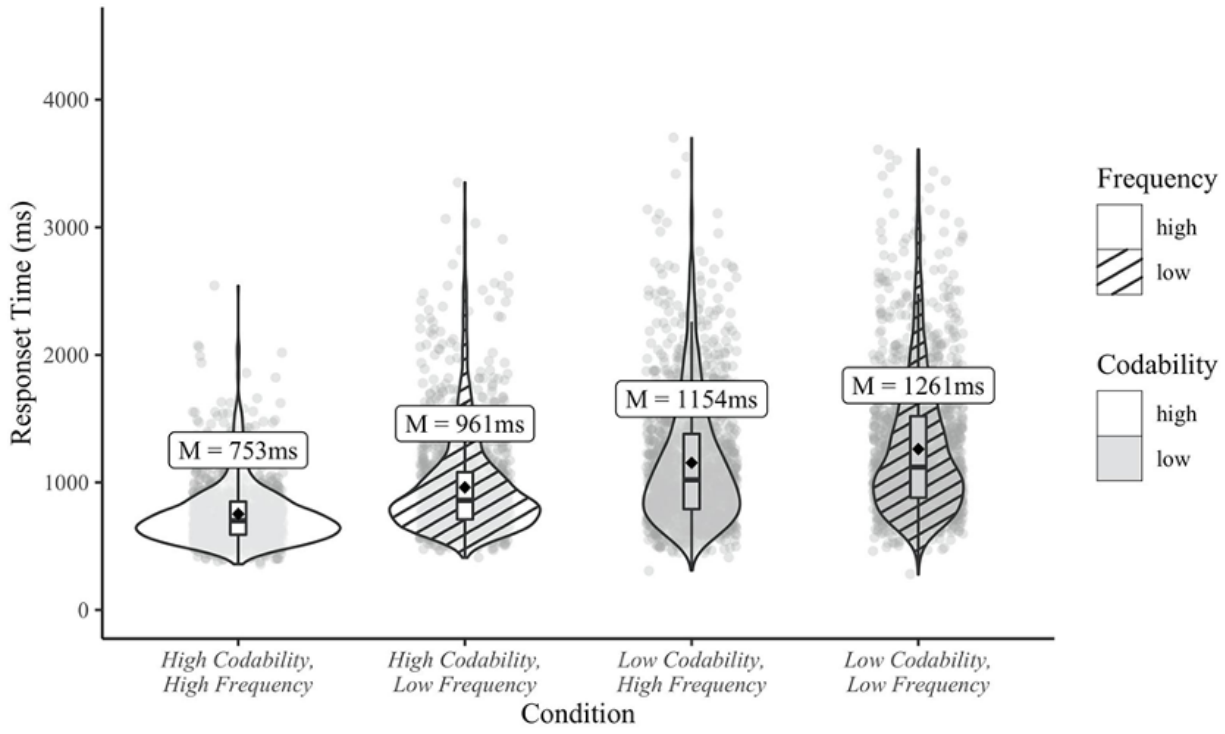
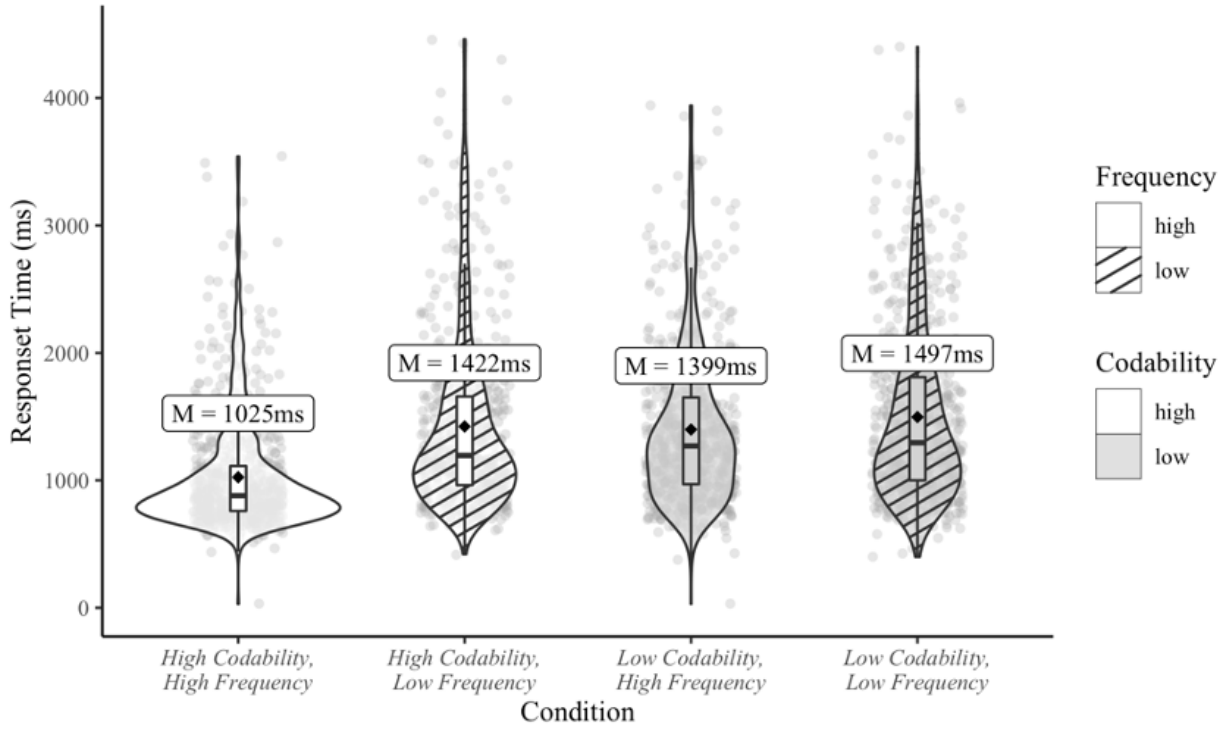


Figure 2.2: *Child RTs by condition.* Each point represents a response. The black diamonds indicate mean RT.



2.2.2.2.2. Adult results

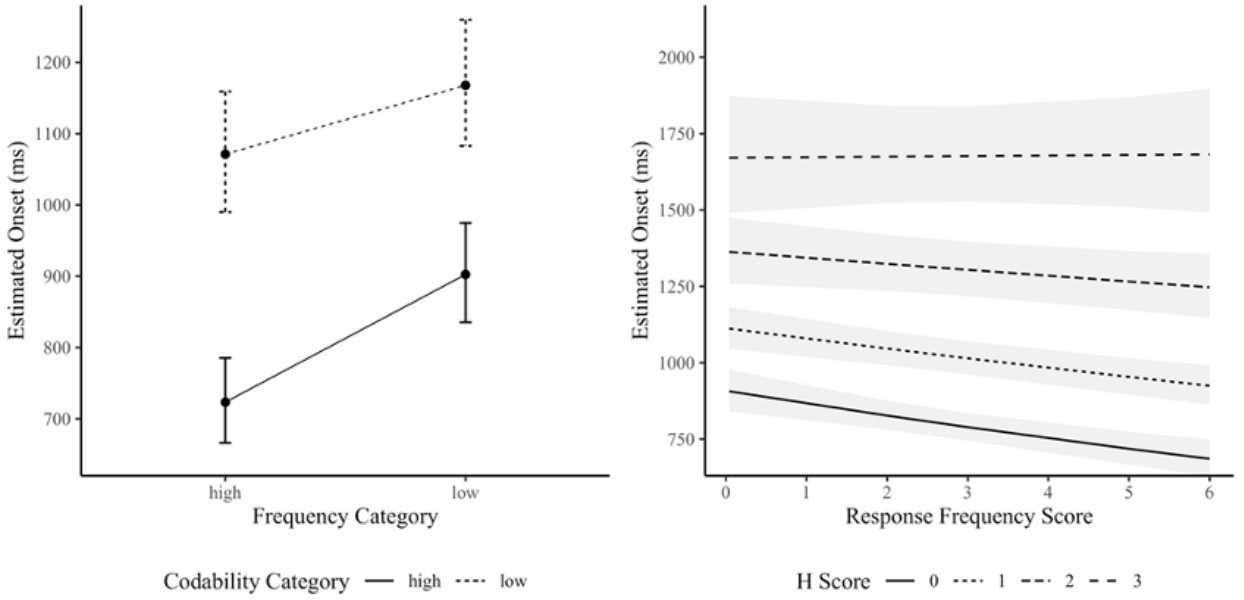
The categorical analysis of the adult data revealed a significant overall main effect of codability category (β = -0.16, $t(126)$ = -10.30, $p < 0.0001$), with longer onset latencies for the low codability items. The estimated RT for low codability items (back-transformed from the log scale) was 1119 ms (95% Confidence Interval [CI] [1050, 1192]), and the estimated RT for high codability items was 809ms (95% CI [759, 861]). We also observed a significant effect of frequency category (β = -0.08, $t(134)$ = -4.32, $p < 0.0001$), with longer latencies for the low frequency items. The estimated RT for low frequency items was 1028 ms (95% CI [965, 1095]), and the estimated RT for high frequency items was 882 ms (95% CI [824, 944]). Effect plots showing the marginal effects of codability and frequency in the adult and child data are available

in the Supplementary Materials. The main effect of target name syllable count was not significant in the present analysis (β = 0.01, t(115) = 0.46, p = 0.65), though the effect was significant in the full set of adult responses (β = 0.04, t(5775) = 5.63, p < 0.0001), with longer RTs for longer target names. The main effect of trial number was significant (β = 2.4e-04, t(5491) = 3.86, p = 0.0001), with longer RTs for later trials.

Crucially, there was a significant interaction between codability category and frequency category (β = -0.03, t(119) = -2.17, p = 0.03) such that the difference between high and low frequency was smaller for the low codability items than the high codability items (Figure 2.3). The effect of frequency category was significant in the high codability categories (β = 0.22, t(134) = 4.47, p < 0.0001) but not the low codability categories (β = 0.09, t(120) = 1.92, p = 0.06).

We observed the same pattern of results in the continuous analysis as in the categorical analysis (see Supplementary Materials for complete result summary). There were significant effects of item adult H score (β = 0.20, t(193) = 8.51, p < 0.0001) and response frequency score (β = -0.05, t(269) = -4.63, p < 0.0001), with slower RTs when H score was higher (i.e., indicating lower codability) and when frequency score was lower. Crucially, we also observed a significant interaction between adult H score and frequency score (β = 0.02, t(205) = 2.59, p = 0.01) that parallels the interaction between codability and frequency category observed in the categorical analysis: As H score increased (i.e., codability decreased), so did the effect of frequency score, meaning that the frequency effect was smaller for items with less name agreement (Figure 2.3).

Figure 2.3: *Interactions between codability and frequency measures in the adult data.* RT estimates have been back-transformed from the analysis scale to the response scale (ms). Error bars and ribbons indicate 95% confidence intervals.



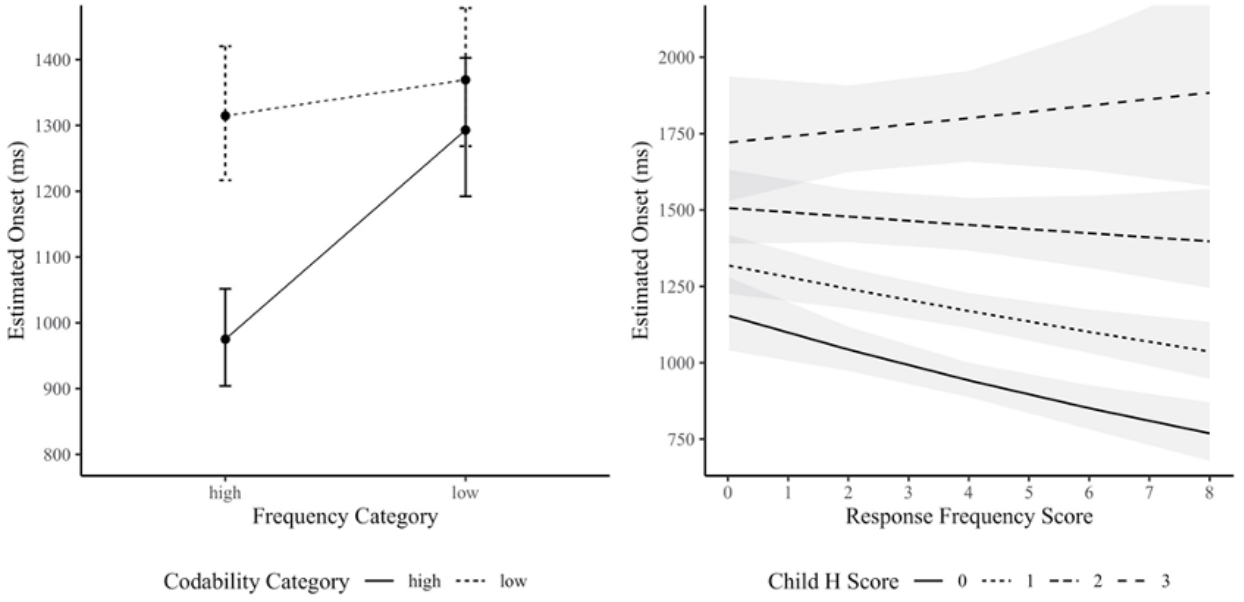
2.2.2.2.3. Child results

The categorical analysis of the child data revealed significant overall main effects of codability category (β = -0.09, $t(86)$ = -5.24, $p < 0.0001$) and frequency category (β = -0.08, $t(99)$ = -4.63, $p < 0.0001$), with significantly longer RTs for the low codability and low frequency items. The estimated RT for low codability items was 1340 ms (95% CI [1258, 1427]), and the estimated RT for high codability items was 1113ms (95% CI [1046, 1184]). The estimated RT for low frequency items was 1330 ms (95% CI [1247, 1418]), and the estimated RT for high frequency items was 1128ms (95% CI [1060, 1201]). The main effects of trial number (β = 9.6e-04, $t(2293)$ = 4.47, $p < 0.0001$) and target name syllable count (β = 0.06, $t(113)$ = 2.18, p = 0.03) were significant, with longer RTs for later trials and for items with longer target names.

Critically, as in the analysis of the adult data, there was a significant interaction between codability category and frequency category (β = -0.06, $t(105)$ = -3.83, $p < 0.001$) such that the effect of frequency was larger when codability was high (Figure 2.4). The effect of frequency category was significant in the high codability categories (β = 0.28, $t(96)$ = 5.80, $p < 0.0001$) but not in the low codability categories (β = 0.04, $t(109)$ = 0.90, p = 0.37).

The same general pattern was present in the continuous analyses (complete results available in the Supplementary Materials). For the analysis using child H score as our continuous measure of codability, we had to drop the interaction term from the participant random effect for model convergence. In the resulting model, we observed significant effects of child H score (β = 0.13, $t(182)$ = 4.58, $p < 0.0001$) and response frequency score (β = -0.05, $t(128)$ = -4.08, $p < 0.0001$), and a significant interaction between the two variables (β = 0.02, $t(281)$ = 2.68, $p < 0.01$) (Figure 2.4). In the analysis using adult H score rather than the child H score, we were able to use the full random effects structure. In this model, we observed a significant interaction between adult H score and response frequency score (β = 0.02, $t(56)$ = 2.04, $p < 0.05$) and a significant effect of response frequency score (β = -0.06, $t(160)$ = -4.59, $p < 0.0001$). The marginal effect of adult H score did not reach conventional levels of significance with the full random effects structure (β = 0.08, $t(65)$ = 1.97, p = 0.05) but did in a model with the same random effects structure as the child H score analysis (β = 0.08, $t(262)$ = 2.09, p = 0.04).

Figure 2.4: *Interactions between codability and frequency measures in the child data.* RT estimates have been back-transformed from the analysis scale to the response scale (ms). Error bars and ribbons indicate 95% confidence intervals.



2.2.2.2.3. Q1 summary

In the Q1 analyses, we replicated previously-observed codability and frequency effects in both adults and five-year-old children: Naming RT was faster when images had high codability (more name agreement) and when their names were more frequent. In both populations, we additionally observed under-additive interactions between the effects of codability and frequency such that the frequency effect was attenuated when codability was low. These effects were observed using both categorical and continuous measures of codability and frequency. We address a possible explanation for these under-additive interactions in Study 3, suggesting that the interaction stems from the dynamics of the language production system.

One concern that is always present in reaction time studies is that timing differences across conditions could be a side effect of differences in the proportion of responses that were discarded. For example, we omitted responses that were not completed within the response window; if these omitted responses were particularly common for items with low codability and low frequency (especially compared to high codability, low frequency items), then the absence of these longer response times could result in an artificial under-additive interaction. We saw no evidence for this pattern. Less than 0.5% of all responses were omitted for this reason in both the adult and child data sets, and these omissions were scattered across the four cells of the design (see Supplementary Materials). Similarly, speech errors resulting in response omission do not appear to be more common for responses to *Low Codability, Low Frequency* items compared to those to *High Codability, Low Frequency* items in either population (see Supplementary Materials). We consequently do not believe that the observed under-additive interactions are a by-product of response omissions.

In addition to observing effects of codability and frequency, our analyses also reproduced other established naming RT patterns in adults and children. Both populations displayed fatigue effects (e.g., D'Amico et al., 2001), with slower response times to later trials, though the fatigue effect was more pronounced in the child data. We also replicated effects of word length (e.g., Bates et al., 2003; D'Amico et al., 2001; Johnson et al., 1996; Székely et al., 2003; Székely et al., 2005), with faster RTs for shorter names. In a post-hoc exploratory analysis, we additionally assessed the influence of age of acquisition (AoA) on naming response time (see Supplementary Materials). Words rated as having been acquired earlier in life tend to be produced faster and with fewer errors (e.g., Carroll & White, 1973; see Brysbaert & Ghyselinck, 2007 for review; see Anderson, 2008; D'Amico et al., 2001; Johnson & Clark, 1988 for evidence of AoA effects in

children). AoA was omitted from our preregistered analysis for two reasons. First, there is a strong correlation between frequency and AoA since more frequent words tend to be learned at earlier ages (see, e.g., Goodman et al., 2008), raising statistical concerns. Second, there is debate over whether AoA measures capture the same cognitive construct as word frequency measures (e.g., Zevin & Seidenberg, 2002). Our post-hoc analyses replicate previously-observed AoA effects in adults and children and provide evidence that AoA and frequency effects pattern differently in our data set, suggesting that frequency and AoA have independent effects (Brysbaert & Ghyselinck, 2007; Juhasz, 2005) and that the observed interaction between codability and frequency is not driven by AoA. In an additional exploratory analysis, we also confirm that the observed interaction between codability and frequency is independent from effects of phonological neighborhood density (see Supplementary Materials).

Overall, we observed very similar results in both the adult and child analyses, suggesting similar underlying processes in both populations. We also found broadly similar results regardless of whether we based our predictors on the child data or the adult data. Nevertheless, in the continuous analysis, we observed a greater estimated effect on child naming RT for child H score (Cohen's $d = 0.68$) than adult H score (Cohen's $d = 0.49$ with the full effects structure, Cohen's $d = 0.26$ with the same effects structure as the child H model) (standardized effects sizes computed using {EMAtools} v0.1.4; Kleiman, 2021). These differences suggest that child H scores may serve as a more accurate reflection of child naming behavior. Differences between adult and child H scores could reflect differences in the words that children know or in the contexts in which they use them, resulting in differences in adult and child name agreement.

The results of the Q1 analyses demonstrate that by five years of age, lexical production in children depends upon both the codability of the referent and the frequency of the word. These

RT effects suggest that the process of lexical production is similar in both adults and five-year-olds, as both populations display codability and frequency effects on picture naming RT. However, mean differences in response time can arise through a variety of different changes in the response time distribution. These changes are thought to reflect different types of processing costs (some of which affect all trials, some of which affect only a subset). In the next section, we explore how our two factors (codability and frequency) affect the response time distribution in children and adults to assess whether the underlying processes affected by these manipulations are qualitatively similar in the two populations.

2.2.3. Q2: Influences of codability and frequency on the response time distribution

In this section, we explore how the codability and frequency manipulations affected the RT distributions in the adult and child data. A variable can affect RT by shifting the mean (increasing the RT for all trials), changing the standard deviation (increasing the variability of RT), or skewing the distribution (increasing the RT for a subset of trials). These different changes are thought to reflect different underlying processes (Balota & Spieler, 1999). By analyzing the RT distribution and how it changed for each manipulation in each population, we can gain insight into the processes involved in lexical production and how they change (or stay the same) between five years of age and adulthood.

2.2.3.1. Q2 data analysis procedure

To investigate how the frequency and codability manipulations impacted the RT distributions in our data, we fit ex-Gaussian distributions (Ratcliff, 1979) to the RT data from each participant in each codability and frequency category (high, low). Ex-Gaussian distributions

are often used to describe RT distributions (Balota & Spieler, 1999; Dawson, 1988; Luce, 1986; Ratcliff, 1993). Ex-Gaussian distributions are convolutions of a normal distribution (described by parameters μ and σ) with an exponential distribution (described by the parameter τ). The way that a manipulation influences these parameters can serve as an indication of how it influences response time. If a manipulation shifts the mean of the RT distribution, it will primarily influence μ. Changes in the standard deviation of the distribution will influence σ. If a manipulation increases the skew of the RT distribution, it will primarily influence τ.

Parameter estimations in our analysis were computed using the maximum likelihood method. We fit the ex-Gaussian distributions in R using functions from the {retimes} v0.1-2 package (Massidda, 2013). We constructed linear mixed effects models (using {lmerTest} v3.1-3; Kuznetsova et al., 2017) to analyze the estimates obtained for each parameter in each category. The models had fixed effects of category level (high vs. low) and manipulation (codability, frequency) with an interaction, as well as a random intercept for participant. We used dummy-coding to establish the contrasts of interest.

We supported this analysis with vincentile plots (Vincent, 1912), which serve as a non-parametric confirmation of ex-Gaussian analyses (e.g., Balota et al., 2008; Staub, 2010). We constructed these plots by dividing the data for each participant in each condition into ten vincentiles (the fastest 10% of responses, the next fastest 10%, etc.). We then calculated the mean RT for each participant in each condition (high, low) in each vincentile as well as the difference between these means. The vincentile plots show the mean RT difference between conditions across all participants at each vincentile; a vincentile plot thus demonstrates how the size of an effect changes across the RT distribution. The shapes of vincentile plots systematically reflect effects on ex-Gaussian parameters (Balota et al., 2008). A manipulation that shifts a RT

distribution, resulting in a change in μ, will have a relatively flat vincentile plot, with similar mean RT differences for all ten vincentiles. Changing the size of σ will leverage the plot around a midpoint, increasing slope as σ increases. A manipulation that skews the RT distribution, manifesting in a change in τ, will have greater mean RT differences in the tail of the distribution (i.e., at larger vincentiles), resulting in a vincentile plot that curves upwards as vincentile number increases.

We performed separate analyses for the adult and child data. For the adult data, we used the same codability and frequency categories as in the Q1 categorical analysis. For the child data, we reclassified the items' codability categories based on their child H scores, because, as we noted in the *Q1 summary*, the child H scores appeared to better capture the codability effect in our child data than adult H scores did. The items were categorized using the same method described in the *Q1 analysis procedure* but with child H score ratios. We refer to these new categories as the "child codability" categories.

Table 2.2 illustrates the adult and child codability and frequency measures for the categories in the analysis. As mentioned in the *Q1 analysis procedure*, the high and low frequency categories and the codability categories based on adult H score only vary along the dimension we intended to manipulate. The child codability categories, on the other hand, were less well-balanced. The high and low child codability items differed significantly based on both adult H score ($t(118) = -7.20$, $p < 0.0001$) and child H score ($t(118) = -12.50$, $p < 0.0001$), though they also differed significantly based on frequency score ($t(118) = 3.90$, $p < 0.001$). The high and low frequency items additionally differed based on child H score ($t(118) = -3.93$, $p = 0.0001$). These imbalances reflect the trend that name agreement tends to be higher for items with high frequency names (e.g., Bates et al., 2003). Given the imbalances in the child codability

categories, we also analyzed the child data using the adult codability categories (relevant differences will be discussed in the text; see Supplementary Materials for full results). We address potential influences of these imbalances in the *Q2 summary*.

Table 2.2: *Properties of the codability and frequency categories.* Mean adult H score, child H score, and frequency score for the high and low codability and frequency categories used in the Q2 analyses. SD in parentheses.

Manipulation	Category	Adult H Score	Child H Score	Frequency Score
Codability (Adult)	high (n=59)	0.07 (0.11)	0.63 (0.75)	2.98 (1.60)
	low (n=61)	1.39 (0.57)	1.64 (0.66)	2.87 (1.25)
Codability (Child)	high (n=33)	0.04 (0.09)	0.08 (0.12)	3.70 (1.57)
	low (n=87)	1.00 (0.77)	1.54 (0.66)	2.63 (1.26)
Frequency	high (n=59)	0.65 (0.76)	0.84 (0.82)	4.14 (0.90)
	low (n=61)	0.82 (0.80)	1.43 (0.82)	1.74 (0.63)

2.2.3.2. Q2 results

The mean estimated values of the three ex-Gaussian parameters for adult and child data in each level of the codability and frequency manipulations are given in Table 2.3 along with the estimated RT differences between the levels. Reliability of the low – high difference for the ex-Gaussian parameters is indicated in the table.

2.2.3.2.1. Adult results

The μ estimates were larger in the low codability condition than the high codability condition ($\beta = 125.82$, $t(141) = 12.34$, $p < 0.0001$) and in the low frequency condition than the

high frequency condition (β = 86.96, t(141) = 8.53, p < 0.0001). There was a significant interaction between level and manipulation such that the difference between the codability conditions was greater than the difference between the frequency conditions (β = -38.86, t(141) = -2.69, p < 0.01).

The σ estimates were larger in the low codability condition than the high codability condition (β = 43.37, t(141) = 5.15, p < 0.0001). There was no reliable difference between the low and high frequency groups in the present analysis (β = 12.07, t(141) = 1.43, p = 0.15), though estimates were reliably larger in the low frequency condition in the full adult data set (β = 16.26, t(141) = 2.12, p = 0.04). There was a significant interaction between level and manipulation such that the codability manipulation had a greater effect on σ than the frequency manipulation (β = -31.30, t(141) = -2.63, p < 0.01).

The τ estimates were larger in the low codability condition than the high codability condition (β = 231.26, t(141) = 18.56, p < 0.0001) and in the low frequency condition than the high frequency condition (β = 83.04, t(141) = 6.67, p < 0.0001). There was a significant interaction between level and manipulation such that the effect on τ was larger for the codability manipulation than the frequency manipulation (β = -148.23, t(141) = -8.41, p < 0.0001).

Figure 2.5 provides the vincentile plots of the codability and frequency manipulations in the adult data. The RT difference between high and low codability conditions increased from the faster to the slower RT vincentiles. The upward curve of the codability plot resembles the idealized plot for a RT effect due to a change in τ (exponential contribution) (Balota et al., 2008). The plotted line for the frequency manipulation, on the other hand, has a more linear shape, with the effect size increasing linearly across the RT distribution. This vincentile plot shape is

consistent with those for RT effects that are divided between μ (mean) and τ (exponential contribution) (Balota et al., 2008).

2.2.3.2.2. Child results

In the child analysis, the μ estimates were larger in the low codability condition than the high codability condition (β = 111.24, $t(72) = 7.71$, $p < 0.0001$) and in the low frequency condition than the high frequency condition (β = 99.12, $t(72) = 6.87$, $p < 0.0001$). There was no significant interaction between level and manipulation for μ (β = -12.13, $t(72) = -0.59$, $p = 0.55$).

The σ estimates were larger in the low codability condition than the high codability condition (β = 68.15, $t(72) = 3.80$, $p < 0.001$). There was no reliable effect of frequency condition on the σ estimates (β = 15.54, $t(72) = 0.87$, $p = 0.39$). The interaction between level and manipulation was significant such that the codability manipulation had a greater effect on σ than the frequency manipulation (β = -52.61, $t(72) = -2.08$, $p = 0.04$), though this interaction did not reach conventional levels of significance in the analysis with adult codability categories (β = -51.69, $t(72) = -1.97$, $p = 0.05$).

The τ estimates were larger in the low codability condition than the high codability condition (β = 288.02, $t(72) = 10.81$, $p < 0.0001$) and in the low frequency condition than the high frequency condition (β = 161.08, $t(72) = 6.05$, $p < 0.0001$). The interaction between level and manipulation was significant such that the effect on τ was larger between the codability conditions than the frequency conditions (β = -126.94, $t(72) = -3.37$, $p = 0.001$). This interaction was not significant in the analysis using adult codability categories.

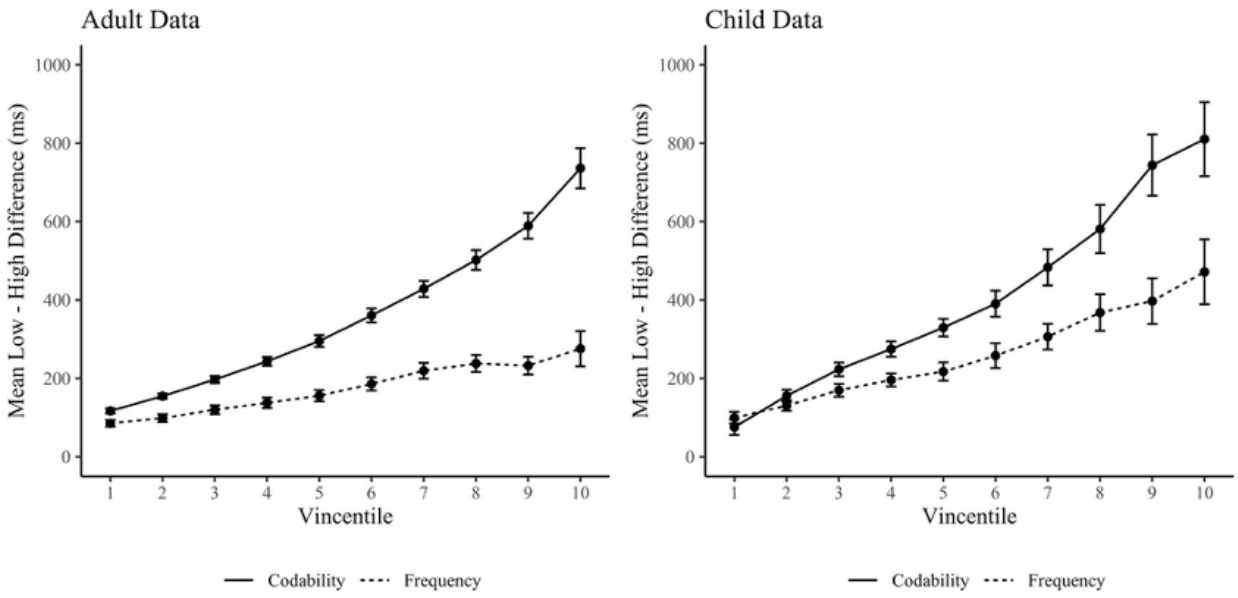
Figure 2.5 shows the codability and frequency vincentile plots for the child data. The codability plot shows a similar upward curve to that observed in the adult analysis, consistent

with a large τ effect (skewing the RT distribution). The plotted line for the frequency manipulation is primarily linear, though it has a slight upward curve; this is consistent with the finding that the τ effect (exponential contribution) for frequency was larger than the μ effect (mean) though still smaller than the τ effect for child codability. The greater slope of the frequency plot compared to the adult frequency vincentile plot is consistent with the fact that children had a larger frequency RT effect (268 ms) than adults (169 ms).

Table 2.3: *Mean RT and ex-Gaussian parameter estimates (in ms) for the high and low codability and frequency conditions*. For the low - high difference reliability, '***' indicates p-values ≤ 0.001, and 'n.s.' indicates p-values > 0.1.

	Adult Data				Child Data			
Condition	RT	μ	σ	τ	RT	μ	σ	τ
High Codability	854	603	65	252	1035	688	53	353
Low Codability	1210	729	108	483	1450	800	121	641
Low - High	356	126	43	231	415	112	63	288
Reliability		***	***	***		***	***	***
High Frequency	948	585	56	364	1194	693	79	500
Low Frequency	1117	672	68	447	1462	792	94	661
Low - High	169	87	12	83	268	99	15	161
Reliability		***	n.s.	***		***	n.s.	***

Figure 2.5: *Vincentile plots.* These plots illustrate the difference between the high and low conditions of each manipulation across the RT distribution. Error bars indicate standard errors.



2.2.3.3. Q2 summary

In the Q2 analyses, we assessed whether the Study 1 codability and frequency manipulations had qualitatively similar influences on adult and child naming RT, suggesting similar underlying processes. We used ex-Gaussian analyses to estimate how codability and frequency influenced the RT distributions for adults and children, investigating which parameters of the distribution were affected by the manipulations.

The ex-Gaussian analyses suggest that codability and frequency effects are qualitatively different from each other but similar across the two age groups. For both children and adults, the frequency manipulation had effects on the μ (mean) and τ (exponential contribution) parameters, suggesting that decreasing frequency both shifts and skews the RT distribution. The codability manipulation, on the other hand, influenced all three parameters in both populations, with a

particularly prominent τ (skew) effect.[7] This skew effect was not only greater than codability's influence on the other two parameters, but it was also considerably larger than the τ effect of the frequency manipulation.

The fact that the codability and frequency manipulations produced different effects on RT distribution suggests that the two variables play different roles in lexical processing. These findings are therefore consistent with hypotheses in which each factor affects a different part of the lexical access process. The ex-Gaussian parameters influenced by codability and frequency manipulations provide clues as to what these processes may be. Differences in the Gaussian parameters (μ and σ) are often interpreted as reflecting changes in automatic processes (Balota & Spieler, 1999), such as the initial perceptual processing or activation of candidates. In contrast, a large shift in the exponential parameter (τ) is the hallmark of decision-making processes (Hohle, 1965). The large τ effect for the codability manipulation thus supports the interpretation of the codability effect as reflecting increased time to resolve competition and select between name candidates during lexical selection. This hypothesis can explain codability's influence on all three parameters. We should expect variation in RT slowdowns for low codability items, as there is variation in the number of candidate names available for each picture (including due to individual differences; e.g., one individual may always refer to a couch using the word *couch*, whereas another may use the words *sofa* and *couch* interchangeably) as well as in the relative frequencies of these names. These variations should result in larger penalties for some trials (e.g., when there are a greater number of names under consideration) and smaller or no influence for

[7] The large τ effect in the child data is unlikely to be solely attributable to the imbalance in target name frequency between the child codability conditions (Table 2.3). The adult data suggest that an effect of frequency should influence τ and μ roughly evenly, meaning that the lower frequency of the low child codability condition should not have disproportionately inflated the τ effect above the μ effect.

others (e.g., when there are fewer name candidates under consideration or one candidate with higher frequency than the others). By contrast, we expect more similar (and faster) RTs in the high codability condition when there are consistently few or no name alternatives to decide between and participants largely produce the same responses. This difference would lead to increased skewing in the low codability condition (a τ effect), and the σ and μ effects would arise as a consequence of these same forces: The variability in the low codability RTs in the low codability condition would increase the standard deviation around the mean relative to the high codability condition (leading to a σ effect), and the larger proportion of slow RTs in the low codability condition would result in a larger mean RT, shifting the distribution (resulting in a μ effect). The fact that codability manipulation influenced all three parameters could additionally reflect influences of the manipulation at multiple levels of processing (e.g., conceptual processing/image identification in addition to lexical selection).

On the other hand, a phonological frequency effect neatly accounts for the RT distribution changes we observed for the frequency manipulation. If the frequency effect reflects differences in the time to activate phonological forms, we should expect similar RT penalties for names with similar frequencies. Thus, we would expect a rightward shift of the RT distribution in the low frequency condition (a μ effect), as speakers are consistently slower to produce these names. Consistent with this account, the frequency manipulation produced a smaller skewing effect (τ effect) than the codability manipulation and no σ effect, demonstrating that slowing in the low frequency condition was more homogenous than in the low codability condition.

While frequency's τ effect was smaller than that for codability, it was still reliable. Skewing of the RT distribution in the low frequency condition could arise for several reasons, all of which are compatible with a phonological frequency effect: (i) Responses were sorted into

high and low frequency categories based on the frequencies of the item dominant names, rather than the individual responses; this leaves open the possibility that these categories do not accurately reflect the frequencies of all the responses they contain. (ii) There is variation in the dominant name frequencies of the items categorized as low codability, which will result in greater RT penalties for some responses compared to others in the category. (iii) There is individual variation in the frequencies with which names are produced and/or encountered, and this variation is likely to be greater for low frequency words than for high frequency words. (iv) Increased skewing for low frequency items may arise due to lateral inhibition at the phonological level or due to the shape of the activation function and different activation thresholds required for selection (Andrews & Heathcote, 2001; Balota & Spieler, 1999). (v) Since item name agreement tends to pattern with name frequency (e.g., naming disparity is smaller for items eliciting high frequency names; Bates et al., 2003), the frequency manipulation may also be a weak manipulation of codability, which could contribute to the skewing effect.

In sum, although the ex-Gaussian analyses do not uniquely support a model of language production in which codability affects lexical selection and frequency affects phonological encoding, they are consistent with it. Minimally, these analyses show that codability and frequency manipulations influence RT distributions in ways that are different from each other.

Critically, the pattern of effects observed for the two manipulations was qualitatively similar in the two age groups investigated. This close parallelism between the children and adults suggests that the mature and developing lexical access systems involve similar underlying mechanisms that exhibit comparable responses to these factors. Nevertheless, there were some more subtle differences between the adult and child groups. The child population demonstrated a larger RT effect of the frequency manipulation than the adult population (268 ms vs. 169 ms,

respectively). In adults, the frequency had similar effects on the mean shift (87 ms) and skew (83 ms), while in children the effect of frequency on skew was more pronounced (161 ms compared to a 99 ms shift effect). This difference may be attributable, in whole or part, to a weak correlation between codability and frequency measures that was present in the child data set but not the adult data set. Specifically, the child H scores for items in the low frequency condition were slightly lower than those for the high frequency items (Table 2.2). For this reason, the frequency manipulation in children is likely also a weak manipulation of codability, resulting in a pattern of effects that is intermediate between the adult frequency and codability effects. This manipulation of codability in the child data would result in an increase in the skew effect for frequency in children, relative to adults, which in turn could contribute to the larger average RT penalty for low frequency trials.

2.2.4. Study 1 discussion

In Study 1, we explored the effects of frequency and codability on lexical production in five-year-old children and adults. Our findings confirmed two previously-reported patterns: Participants in both populations were faster to name images with higher codability (higher name agreement) and with higher frequency. In addition, we went beyond the prior work and fit ex-Gaussian distributions to the reaction time data. We found that codability and frequency manipulations have distinct influences on RT distributions, and these signature patterns are present in both the adult and child data. Taken together, these findings suggest that codability and frequency influence different underlying processes and that their effects are similar in the mature and developing lexical production systems, potentially indicative of similar underlying processes.

Critically, Study 1 found that codability and frequency effects interact in both adult and five-year-old naming behavior: In both the adult and child data, the frequency effect was diminished when codability was low. To our knowledge, such an interaction has not been previously reported for either population. Indeed, early studies suggested that these effects are independent (Lachman, 1973; Lachman & Lachman, 1980), making the interactions observed in our data particularly unexpected. The observed interactions are noteworthy, because in standard models of reaction times, an interaction between two factors implies that these two variables influence at least one common process, either because both are inputs into that process or because they are inputs into separate processes that then interact (Sternberg, 1984, 1998). This is potentially surprising because codability and frequency are thought to influence distinct processes within word planning: the processes of lexical selection and phonological encoding, respectively. Thus, an interaction might suggest that these two processes do not occur in a strict sequence but instead influence one another, as is predicted by a cascading activation architecture.

In Study 2, we investigate the reliability of the observed interaction between codability and frequency in a secondary analysis of data from a previous naming study conducted with adults in several languages (Bates et al., 2003). In Study 3, we explore how such an interaction may arise based on the relationship between lexical selection and phonological encoding in the production planning architecture.

2.3. Study 2: A secondary analysis of prior adult picture naming data

The goal of Study 2 was to determine whether the interactions that we observed in Study 1 between codability and frequency could be observed in other data sets with different properties. Study 1 had two clear limitations. First, we explored naming in just one language

(English). If our hypothesis is correct and the informational cascade is a foundational property of the linguistic architecture, then we should see this same under-additive interaction in other languages as well. Second, our experimental stimuli were initially selected to create four conditions that orthogonally manipulated our two variables, supporting a categorical analysis. As a result, the items did not represent a full spectrum of codability or frequency measures. Thus, it is possible that the interaction effect we observed is a side effect of stimulus selection under these constraints or is limited to extreme values of frequency and codability.

To investigate the generality of the interaction, we looked for parallel effects in the data from Bates et al.'s (2003) multi-language timed picture naming study. In this paper, the authors reported influences of name agreement and word frequency on mean RT in the expected directions (higher H scores and lower word frequencies predicted slower mean RTs), but they did not test for an interaction between codability and frequency RT effects. This data set complements ours because the stimuli were not selected specifically with a manipulation of codability and frequency in mind. Furthermore, it allows us to explore whether the interaction we observed is present across a range of languages.

2.3.1. Methods

We re-analyzed the picture naming data from Bates et al.'s (2003) multi-language naming study, using the data set available from https://crl.ucsd.edu/experiments/ipnp/7lgpno.html. Our analysis focused on the data for adult native speakers of: English ($N = 50$), German ($N = 30$), Hungarian ($N = 50$), Italian ($N = 50$), Mandarin Chinese ($N = 50$), and Spanish ($N = 50$).[8] Participants named a set of 520 black-and-

[8] Bates et al. (2003) also elicited data from native speakers of Bulgarian, however we decided to exclude this language from our analysis, as the frequency measure provided was in the form of subjective ratings

white images. For details about the stimulus materials and experiment methods used, please see Bates et al. (2003).

2.3.2. Analysis

The data set available from Bates et al. (2003) provides mean reaction times for each item in each language (rather than individual trial response data). To parallel our RT analyses in Study 1, we conducted linear regressions on log-transformed mean RT (in log milliseconds). We analyzed the data from each language separately using regression models with fixed effects of item H score, dominant name frequency score, and dominant name syllable count, with an interaction between H score and frequency score. We used the dominant name frequency scores and H scores provided in the Bates et al. (2003) data set. The frequency scores in the data set were derived as natural log transformations from words per million corpus counts (as in Study 1). The properties of the codability and frequency measurements for each language are summarized in Table 2.4.

We also analyzed the data from all six languages in a single model with fixed effects of item H score, dominant name frequency score, language, and dominant name syllable count, with a three-way interaction between H score, frequency score, and language.

instead of derived from corpus counts. It is not clear that we should expect a frequency measure on an ordinal scale to show the same interaction with H score as the frequency score measure used in Study 1.

Table 2.4: *H score and frequency score measures for each language in Bates et al. (2003).*

	H score		Frequency score	
Language	Mean	Range	Mean	Range
English	0.67 (SD = 0.61)	0.00–2.90	2.50 (SD = 1.57)	0.00–7.40
German	0.76 (SD = 0.68)	0.00–3.28	2.01 (SD = 1.50)	0.00–6.62
Hungarian	0.91 (SD = 0.73)	0.00–3.52	1.38 (SD = 1.93)	0.00–6.84
Italian	0.95 (SD = 0.73)	0.00–3.47	1.17 (SD = 1.43)	0.00–6.20
Mandarin	1.16 (SD = 0.79)	0.00–3.56	3.05 (SD = 1.65)	0.00–7.60
Spanish	0.86 (SD = 0.72)	0.00–2.90	2.77 (SD = 1.78)	0.00–8.32

2.3.3. Results

The results of our analyses are summarized in Table 2.5. All six languages analyzed had significant fixed effects of H score and frequency score. Mean RTs increased as H scores increased (i.e., codability decreased) and as frequency score decreased. The interaction between H score and frequency score was significant in the English, German, Mandarin, and Spanish data: As H score increased, the effect of frequency score decreased. There was no statistically significant interaction in Hungarian or Italian. In the combined model looking at the data from all languages, we observed significant fixed effects of H score, frequency score, and their interaction (see Supplementary Materials for the complete model summary).

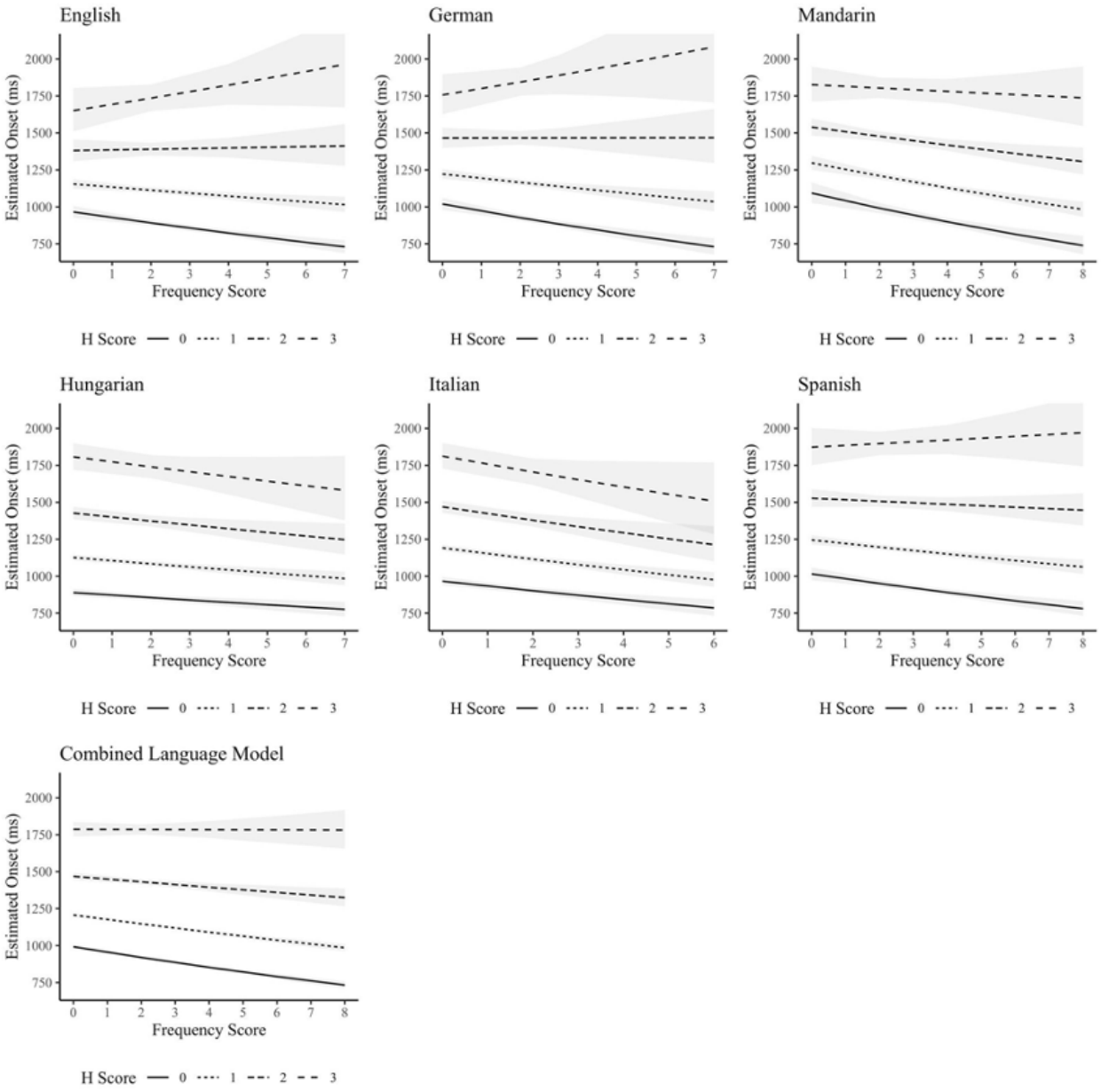
Plots showing the interaction between H score and frequency score in each analysis are presented in Figure 2.6. The plots for the languages with significant interactions between H score and frequency score (English, German, Mandarin, and Spanish) resemble those for the adult and child data in Study 1 (Figures 2.3 and 2.4): The slope of the estimated frequency effect becomes less negative at higher H score values (i.e., at lower levels of name agreement). We observe a

similar pattern in the interaction plot for the combined language analysis. For the two languages without significant interactions (Hungarian and Italian), the slope of the estimated frequency effect is similar at each H score level.

Table 2.5: *Model results for the re-analysis of Bates et al.'s (2003) multi-language naming data.*

Language	H score	Frequency score	Syllable count	Interaction
English	$\beta = 0.18$ $t(515) = 9.11$ $p < 0.0001$	$\beta = -0.04$ $t(515) = -5.97$ $p < 0.0001$	$\beta = 2.7e\text{-}03$ $t(515) = 0.30$ $p = 0.76$	$\beta = 0.02$ $t(515) = 3.19$ $p < 0.01$
German	$\beta = 0.18$ $t(515) = 10.61$ $p < 0.0001$	$\beta = -0.05$ $t(515) = -6.25$ $p < 0.0001$	$\beta = 0.01$ $t(515) = 0.85$ $p = 0.40$	$\beta = 0.02$ $t(515) = 3.22$ $p = 0.001$
Hungarian	$\beta = 0.24$ $t(515) = 20.57$ $p < 0.0001$	$\beta = -0.02$ $t(515) = -3.46$ $p < 0.001$	$\beta = 2.2e\text{-}03$ $t(515) = 0.29$ $p = 0.77$	$\beta = 1.1e\text{-}04$ $t(515) = 0.02$ $p = 0.98$
Italian	$\beta = 0.21$ $t(515) = 18.20$ $p < 0.0001$	$\beta = -0.03$ $t(515) = -4.75$ $p < 0.0001$	$\beta = -0.01$ $t(515) = -1.04$ $p = 0.30$	$\beta = 1.3e\text{-}03$ $t(515) = 0.19$ $p = 0.85$
Mandarin	$\beta = 0.17$ $t(515) = 9.09$ $p < 0.0001$	$\beta = -0.05$ $t(515) = -5.61$ $p < 0.0001$	$\beta = 0.02$ $t(515) = 1.47$ $p = 0.14$	$\beta = 0.01$ $t(515) = 2.62$ $p < 0.01$
Spanish	$\beta = 0.20$ $t(515) = 12.58$ $p < 0.0001$	$\beta = -0.03$ $t(515) = -5.28$ $p < 0.0001$	$\beta = -8.1\text{-}04$ $t(515) = -0.11$ $p = 0.91$	$\beta = 0.01$ $t(515) = 2.69$ $p < 0.01$
Combined Language Model	$\beta = 0.18$ $t(3095) = 8.65$ $p < 0.0001$	$\beta = -0.04$ $t(3095) = -5.91$ $p < 0.0001$	$\beta = 1.7e\text{-}03$ $t(3095) = 0.48$ $p = 0.63$	$\beta = 0.02$ $t(3095) = 3.03$ $p < 0.01$

Figure 2.6: *Codability (H score) × frequency interaction plots for the reanalysis of the Bates et al. (2003) data.* RT estimates have been back-transformed from the analysis scale to the response scale (ms). Ribbons indicate 95% confidence intervals.



2.3.4. Study 2 discussion

Study 2 demonstrates that the under-additive interaction between codability and frequency observed in Study 1 generalizes across several languages and to a different stimulus set. We observed interactions between codability and frequency measures in the same language investigated in Study 1 (English) as well as in three additional languages (German, Mandarin Chinese, and Spanish). These interactions followed the same pattern as the interaction effects observed in Study 1: The frequency effect attenuated as codability decreased. We also observed a similar pattern when combining the data from multiple languages.

We did not observe a reliable under-additive interaction in all languages investigated, however: There were two languages (Hungarian and Italian) in which there was no reliable interaction. While it is possible that these cross-linguistic differences may be attributable to properties of the languages themselves or to different processing strategies in different languages, we believe that the lack of an interaction is most likely due to the frequency measures used for these languages (see discussion in Bates et al., 2003). Specifically, the Hungarian and Italian frequency measures were each derived from a corpus of approximately 500,000 words, which is much smaller than is typically used in psycholinguistic research (Füredi & Kelemen, 1989 for Hungarian; De Mauro et al., 1993 for Italian). In contrast, the frequency measures in the languages that displayed an interaction were based on corpora of approximately 2 million words or greater (Baayen et al., 1995 for English and German; Chinese Knowledge Information Processing Group, 1997 for Mandarin; Alameda & Cuetos, 1995 for Spanish). In addition, the mean frequency scores for Hungarian and Italian were lower than for the languages that displayed an interaction (Table 2.4). A less representative frequency measure with fewer

observations in the high frequency range may make it more difficult to detect variations in the size of the frequency effect at different levels of name agreement.

Another property of the Bates et al. (2003) data set that may have influenced our ability to detect interactions between codability and frequency effects is that the data set provides mean RT per item rather than RT data for the individual elicited responses. By analyzing summary statistics rather than the measures for the individual responses, we lose some of the granularity in the analysis, which could potentially obscure trends. In fact, the present analyses failed to replicate effects of word length found in other studies (e.g., D'Amico et al., 2001; Johnson et al., 1996; Székely et al., 2003; Székely et al., 2005), which supports the hypothesis that the measures used may not be sensitive enough to capture all naming RT trends.

The fact that we observed under-additive interactions between codability and frequency measures in several languages from the Bates et al. (2003) data suggests that the interactions observed in Study 1 were not merely by-products of the particular stimulus set we chose. This is what one would expect if the interaction results from an informational cascade that is a fundamental property of the language production architecture. In Study 3, we further test this hypothesis by simulating how the relationship between lexical selection and phonological encoding processes influences response time.

2.4. Study 3: Simulating the source of the interaction between codability and frequency

In Study 3, we investigated whether an interaction between codability and frequency can arise as a natural consequence of an architecture that allows for cascading activation. We conducted two simulations which predicted naming time based on H score and word frequency. Our simulations focused on estimating how the relationship between lexical selection and

phonological encoding influences naming RT; other processes that affect naming response time, such as conceptual access and articulation planning, are orthogonal to our primary question of how information flows between the lexical selection and phonological encoding stages of word planning. By including only lexical selection and phonological encoding in our simulations, we can get a sense of how the interplay between these processes is able to impact RT. If the relationship between these two processes on its own produces an interaction similar to those observed in Study 1 and Study 2, that provides a proof of concept that the particular architectural choice simulated could underlie the adult and child naming behavior we observed.

We conducted both a serial simulation intended to approximate strictly sequential activation of these two levels of representation as well as a dynamic simulation intended to approximate a simple information cascade between lexical selection and phonological encoding. We analyzed the RTs generated by these simulations to see whether they produced interactions between H score and frequency comparable to those observed in Studies 1 and 2, which would demonstrate that such an interaction can arise from the simulated relationship between lexical selection and phonological encoding. For the purposes of the simulations, we assume that H score influences how long it takes a speaker to select a name for articulation during lexical selection (Alario et al., 2004; Griffin, 2001) and that word frequency score influences the duration of phonological encoding (Griffin & Bock, 1998). Consequently, we used H score and frequency score to estimate the durations of lexical selection and phonological encoding, respectively.

In our serial simulation, the lexical selection and phonological encoding processes were sequential, with phonological encoding taking place only after lexical selection was complete (Figure 2.7). In this simulation, naming RT was equal to the sum of the estimated durations of

the lexical selection and phonological encoding processes. This simulation was intended to approximate a strict discrete serial activation architecture of word planning, in which activation only spreads to phonological form representations after a lexical representation has been selected for articulation.

In our dynamic simulation, phonological encoding of the name to be produced was initiated at the same time as the lexical selection process, approximating a cascading architecture of word planning in which phonological form activation begins shortly after the corresponding lexical representations are first activated (Figure 2.8). In the dynamic simulation, naming RT for a response was determined based on the relative estimated durations of lexical selection and phonological encoding. If the estimated duration of lexical selection was longer than the estimated time to encode the name to be produced, then the RT was equal to the duration of lexical selection (Figure 2.8a). This outcome simulates a situation in which the phonological form of the word ultimately articulated is fully activated via cascading activation even before the speaker has decided which candidate word to produce. If the estimated duration of lexical selection was shorter than the estimated phonological encoding time, then the RT was equal to the duration of phonological encoding (Figure 2.8b). This outcome simulates a situation in which lexical selection is completed before the phonological form of the word to be articulated is fully activated, in which case the word cannot be not produced until encoding of its form is completed. If the estimated durations of lexical selection and phonological encoding were equal, the RT was equal to that duration (Figure 2.8c).

Although these simulations are simplifications of the word planning process (not intended to carefully reconstruct the complexity of real-world language production), if the simplified models can give rise to an under-additive interaction between codability and

frequency, that would provide preliminary evidence that the interactions observed in Studies 1 and 2 can arise as a natural consequence of the corresponding planning architectures.

Figure 2.7: *Schematic showing the relationship between lexical selection time and phonological encoding time on RT in the serial simulation.* The dotted line represents the naming RT of a hypothetical response. The blocks labelled lexical selection and phonological encoding represent the durations of each process for that response. This simulation approximates a strict serial planning architecture in which phonological encoding does not start until after lexical selection is complete.

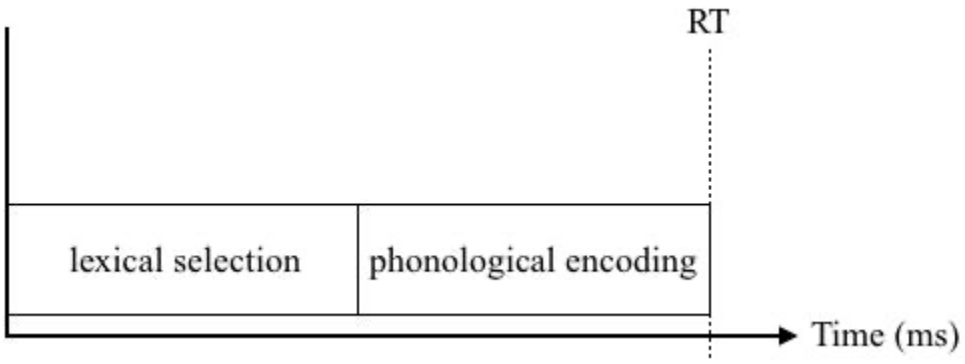
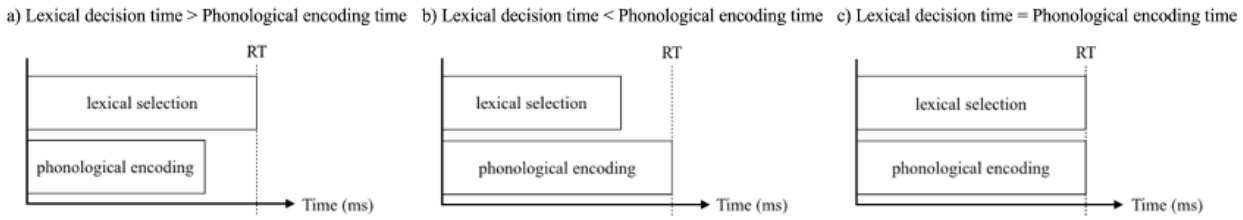


Figure 2.8: *Schematics showing the relationship between lexical selection time and phonological encoding time on RT in the dynamic simulation.* The dotted line represents the RT of the response. The blocks labelled lexical selection and phonological encoding represent the durations of each process for that response. This simulation approximates a planning architecture in which lexical selection and phonological encoding begin simultaneously.



2.4.1. Methods

2.4.1.1. Generating hypothetical responses

The hypothetical responses in our simulations were minimally defined: The only properties attributed to each response were a word frequency score and a referent H score. To ensure that any observed RT effects in the simulations do not result from imbalances in the distribution of these properties within the data set, our hypothetical data sample was constructed in a balanced grid such that the data included samples with all possible combinations of frequency and H score values. Frequency score values ranged from 0.0–10.6 (the range of frequency scores of tokens in the SUBTLEX-US corpus; Brysbaert & New, 2009), and H score values ranged from 0.0–3.6 (the range of H scores in Bates et al.'s, 2003 data set, collapsing across all languages). Both variables were incremented within our grid by 0.1, leading to a sample of 3959 hypothetical responses.

2.4.1.2. Estimating lexical decision and phonological encoding duration

We used the Bates et al. (2003) data set to approximate the marginal effects of H score and frequency score on RT, which we then used to estimate the lexical selection and phonological encoding times (respectively) for our hypothetical responses. We combined the data for all six languages in Bates et al. (2003) included in the Study 2 analysis and constructed a linear model predicting naming RT in log milliseconds. We analyzed the mean RT for dominant name productions in order to maximize the accuracy of the frequency and word length measures in the data set. The model had fixed effects of H score, dominant name frequency score, and language (with a three-way interaction) as well as a fixed effect of dominant name syllable count. We applied the *effect()* function from the {effects} package v.4.2-0 (Fox & Weisberg, 2018) to extract the marginal effects of H score and frequency score from the model. We then used the regression lines of these marginal effects to estimate the influence of H score and frequency score on log RT. The regression equation for H score was $y = 0.1950053x + 6.816174$, and the equation for frequency score was $y = -0.02744063x + 7.047661$.

For each hypothetical response, we entered its H score and frequency score into the corresponding linear equation and exponentiated the resulting values to obtain time estimates in milliseconds. We used these time outputs as our estimates of the duration of the lexical selection and phonological encoding processes. For example, a hypothetical response with an H score of 1.0 and a frequency score of 1.0 would have an estimated lexical selection duration of *exp(0.1950053*1.0 + 6.816174)*, or approximately 1109 ms, and an estimated phonological encoding duration of *exp(−0.02744063*1.0 + 7.047661)*, or approximately 1119 ms.

2.4.1.3. Calculating naming RT

We used the lexical selection and phonological encoding durations for each hypothetical response to determine its naming RT in our simulations. The same set of hypothetical responses was used in both simulations. In these simplified simulations, a response was considered "produced" (i.e., named) as soon as (i) its lexical duration had elapsed (i.e., simulating that the response name had been selected at the lexical level) and (ii) its phonological encoding duration had elapsed (i.e., simulating that the response's phonological form fully retrieved and encoded).

In the serial simulation, the RT of a response was equal to the sum of its lexical selection duration and its estimated phonological encoding duration. For example, if a response had a lexical selection duration of 1200 ms and a phonological encoding duration of 1100 ms, its RT would be 2300 ms in the serial simulation. In the dynamic simulation, if a response's lexical selection duration was longer than its estimated phonological encoding duration, its RT was equal to its lexical selection duration. For the above example, the response's RT would thus be 1200 ms in the dynamic simulation. If a response's estimated lexical selection duration was shorter than its estimated phonological encoding duration, the RT was equal to the phonological encoding duration. For example, for a response with a lexical selection duration of 1200 ms and a phonological encoding duration of 1250 ms, the RT would be 1250 ms. If a response's lexical selection and phonological encoding durations were equal, the RT was equal to that duration.

2.4.2. Analysis

We analyzed the RTs produced by the two simulations for the hypothetical responses using linear regression models. To parallel the RT analyses in Studies 1 and 2, we analyzed log-transformed RTs (log milliseconds). We constructed separate regression models for each

simulation. These models had fixed effects of H score and frequency score with an interaction between them.

2.4.3. Results

The RTs were on average longer in the serial simulation (M = 2323 ms, SD = 287 ms, range = 1772–2991 ms) than in the dynamic simulation (M = 1335 ms, SD = 262 ms, range = 913–1841 ms). This is an expected by-product of the way that the RTs were calculated.

Figure 2.9 shows the relationships between H score, frequency score, and RT predicted by the analysis models for the serial and dynamic simulations. The shape of this relationship was very different in the two simulations. The interaction plot for the serial simulation shows no obvious change to the slope of the frequency effect as H score increases. In contrast, the shape of the interaction plot for the dynamic simulation closely resembles the interactions between H score and frequency observed in Studies 1 and 2; as H score increases, the negative slope of the frequency effect decreases.

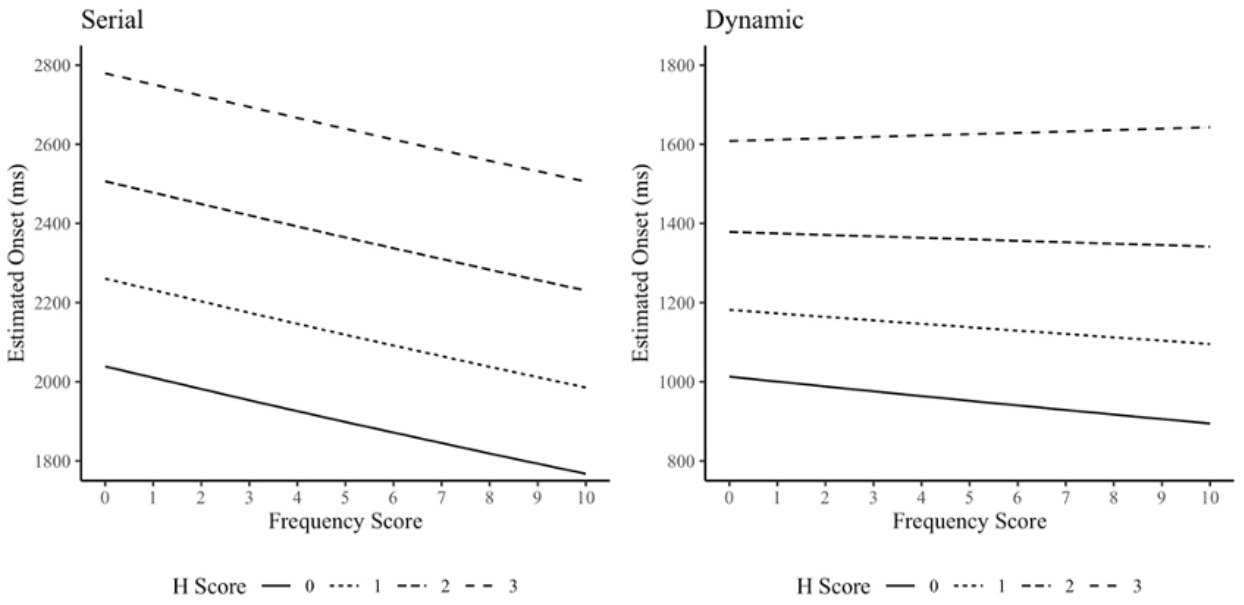
The main effects of H score and frequency score were significant for both the serial simulation (H score: $\beta = 0.10$, $t(3955) = 728.79$, $p < 0.0001$; frequency score: $\beta = -0.01$, $t(3955) = -293.30$, $p < 0.0001$) as well as the dynamic simulation (H score: $\beta = 0.15$, $t(3955) = 215.33$, $p < 0.0001$; frequency score: $\beta = -0.01$, $t(3955) = -50.81$, $p < 0.0001$). RTs were shorter for responses with lower H scores (higher codability) and for responses with higher frequency scores.

The interaction between H score and frequency was significant in the dynamic simulation ($\beta = 4.8e\text{-}03$, $t(3955) = 41.49$, $p < 0.0001$), where the interaction effect plot shows an attenuation of the frequency effect as H score increases (Figure 2.9). Surprisingly, this interaction was also

significant for the serial simulation (β = 1.3e-03, t(3955) = 56.48, p < 0.0001), where there is no obvious change in the slope of the frequency effect by H score value in the interaction plot (Figure 2.9). Further exploration of the data suggests that the interaction in the serial simulation (but not the dynamic simulation) is a by-product of log transforming the dependent variable in the analysis. Repeating our analyses using untransformed RTs, the interaction between H score and frequency score is no longer significant in the serial simulation (β = 0.00, t(3955) = 0.00, p = 1), even though the relationship between H score, frequency score, and RT appears almost identical to that in Figure 2.9 (see Supplementary Materials). By contrast, the pattern of results in the dynamic simulation remains the same when analyzing untransformed RTs (see Supplementary Materials). Importantly, as in the dynamic simulation, we continue to see evidence of under-additive interactions between H score and frequency score measures in the Study 1 data when using untransformed RTs for analysis; we observe interactions in both the adult and child data (though the interaction only reaches conventional levels of significance in the adult data when using a simplified random effects structure; see Supplementary Materials).

To further explore whether the serial simulation produces an under-additive interaction between codability and frequency measures similar to those observed in Study 1, we performed a categorical analysis of the simulated data in the style of the Study 1 categorical analysis (see Supplementary Materials). In this analysis, there was no significant interaction between codability and frequency category in the serial simulation (β = -3.26e-03, t(3955) = -1.12, p = 0.26). In the dynamic simulation, there was a significant under-additive interaction similar to those observed in Study 1, with an attenuated frequency effect when codability is low (β = -0.03, t(3955) = -5.96, p < 0.0001). This pattern of results does not change when using untransformed RT as the dependent variable.

Figure 2.9: *Interactions between H score and frequency score in the Study 3 simulations.* RT estimates have been back-transformed from the analysis scale to the response scale (ms). Ribbons indicate 95% confidence intervals.



2.4.4. Study 3 discussion

In Study 3, we compared the interaction between referent H score and word frequency score in two simulations with different relationships between lexical selection and phonological encoding. In the serial simulation, phonological encoding began only after lexical selection was complete. In the dynamic simulation, phonological encoding began at the same time as lexical selection, simulating a simple informational cascade between representations at the two processing levels.

The dynamic simulation produced response times that exhibited an attenuation of the frequency effect as H score increased (Figure 2.9). This interaction between codability and frequency had a similar shape to those observed in Studies 1 and 2. The response times produced

by the serial simulation, on the other hand, did not show the same under-additivity between codability and frequency (Figure 2.9).

Surprisingly, in our primary analyses, the interaction between H score and frequency score was significant for both simulations. However, we recommend that the interaction in the serial simulation analysis be interpreted with caution. Unlike in the dynamic simulation, there is no comparable interaction for the serial simulation when analyzing the simulated RTs in a categorical fashion. Moreover, the interaction in the continuous analysis of the serial simulation appears to be a by-product of analyzing log-transformed RT. The interaction in the serial simulation does not persist in an analysis of untransformed response time, even though the relationship between H score, frequency score, and RT appears virtually identical in the modelled interaction (see Supplementary Materials for plot). Critically, the interaction does persist when analyzing untransformed RTs in the dynamic simulation as well as the Study 1 data. Nevertheless, even if one assumes that the interaction between H score and frequency score is reliable in the serial simulation, the shape of this interaction (Figure 2.9) is so drastically different from those observed in Studies 1 and 2 that it would be unreasonable to assume that they result from the same underlying relationship between H score, frequency score, and RT.

The fact that a clear and reliable under-additive interaction between codability and frequency is produced by a simulation of word planning in which phonological encoding of the produced word begins before lexical selection has been completed (but not by a simulation when the two processes occur sequentially) supports the hypothesis that the interaction effects observed in Studies 1 and 2 can arise as a natural consequence of cascading activation between lexical and phonological forms during word planning.

It is important to note, however, that the simulations in Study 3 have limitations to their psychological plausibility. We identify several of these in Table 2.6 along with the potential consequences they have for our interpretation of the simulations. We conclude that these simplifications are unlikely to account for the differences between the two simulations and are unlikely to prevent the simulations from capturing the relationship between codability and phonological frequency effects. We draw the reader's attention to the final two of these limitations, which, on the face of it, seem most relevant to our theoretical interpretation of the interaction between codability and frequency.

First, the simulations do not model any competition effects between forms at the phonological encoding level (note that competition at the lexical selection level is modelled by the H score effect). We know that phonological competition does occur during picture naming: Reaction times are longer for names that are in more dense phonological neighborhoods (e.g., Sadat et al., 2014; Zhang et al., 2020). Nevertheless, it seems unlikely that phonological competition would fundamentally change the behavior of the simulations or offer an alternative explanation for the critical interaction.

On the one hand, adding phonological competition to the serial simulation would not, in and of itself, produce an under-additive interaction between codability and frequency, since the effects of codability would necessarily occur before and separate from phonological encoding. In a strict serial architecture, phonological competition would come from phonological neighbors of the target word to be articulated. We would not expect an under-additive interaction unless the names applied to referents with low codability reliably have larger phonological neighborhood densities than those applied to referents with higher codability, leading to differential slowing at different levels of name agreement. This does not appear to be the case: Investigating the

relationship between H score and dominant name phonological neighbor count in Bates et al.'s (2003) English, German, and Spanish data for single word dominant names, we did not observe such a trend (phonological neighborhood size was derived from CLEARPOND; Marian et al., 2012; CLEARPOND data was not available for the other languages in the data set). In the Spanish data, H scores were not significantly correlated with dominant name phonological neighborhood count (Pearson's $r = 0.03$, $t(501) = 0.77$, $p = 0.44$), and there was a significant negative correlation in the English (Pearson's $r = -0.12$, $t(518) = -2.66$, $p < 0.01$) and German (Pearson's $r = -0.08$, $t(518) = -2.02$, $p = 0.04$) data, meaning that phonological neighborhood size was greater for items with higher codability (more name agreement). A negative relationship between phonological neighborhood and name agreement predicts attenuation of the frequency effect at higher levels of codability rather than lower ones, producing the opposite pattern of the codability and frequency interaction we observed in Studies 1 and 2.

On the other hand, adding phonological competition to the dynamic simulation is unlikely to eliminate the under-additive interaction. If multiple phonological forms receive spreading activation from an informational cascade, one might expect increased competition between these activated forms, even after only one lexical representation is selected for articulation. In fact, in Study 1 participants produced more false starts in the low codability conditions than the high codability conditions (see Supplementary Materials), which could reflect an influence of activated word forms other than the one the speaker intended to produce. However, in the dynamic simulation, adding phonological competition is more likely to accentuate the under-additive interaction than to erase it. Specifically, in the dynamic simulation, phonological competition would likely be greater when name agreement is low, allowing for phonological competition from unselected candidates (and their phonological neighbors). This

would lead to slower response times for low codability pictures, particularly when the preferred label was low frequency (relative to the competitors). This slowing may further attenuate any potential processing boost the target has from increased form frequency at higher H score levels, resulting in a greater under-additive effect. Consequently, we do not believe adding phonological competition to either the serial or dynamic simulation would change the pattern of observed results.

A second limitation of the simulations, and the one that seems most relevant to the critical interaction, is that they do not account for frequency effects at the level of lexical selection. As mentioned in the Introduction, there is evidence that word frequency may influence lexical selection processes in addition to the selection of phonological form (e.g., Finocchiaro & Caramazza, 2006; Jescheniak & Levelt, 1994; Johnson et al., 1996; Kittredge et al., 2008; Strijkers et al., 2010). By attributing frequency effects solely to phonological encoding, the simulations may overestimate the size of the phonological frequency effect, though as mentioned in Table 2.6, we are not concerned about capturing the precise magnitude of this effect. More crucially, if both frequency and name agreement influence the process of lexical selection, that creates an opportunity for the two factors to interact at that level. While there is no particular reason to assume that the factors would interact, or that this interaction would be under-additive, this does open the possibility that a serial simulation with these features might produce the critical interaction. We discuss this possibility further in the General discussion.

Despite these simplifications, the simulations provide a rough sketch of how referent name agreement and word frequency may interact in word planning architectures with different relationships between lexical selection and phonological encoding. The simplicity of the simulations allows us to see how the relationship between these two processes can influence

naming RT in the language production architecture under idealized conditions. The simulations show that an under-additive interaction between codability and frequency, similar to that observed in Studies 1 and 2, arises naturally from a planning architecture in which phonological planning begins before lexical selection has been completed, approximating an informational cascade between the two processing levels. Coupled with the prior evidence for cascaded processing in adults (see *§2.1. Introduction*) and the qualitatively similar naming behavior between adults and five-year-olds in Study 1, we believe that these results support the hypothesis that cascading activation is present in the five-year-old language production system.

Table 2.6: *Limitations of the Study 3 simulations.*

Limitation	Potential Consequences
The simulations simplify the word production process to two steps (lexical selection, phonological encoding) and do not take into account other processes involved in naming (e.g., conceptual processing, articulatory planning).	If these other processes are serial and separate, then our simulations capture the naming process with reasonable accuracy (just add constants for each process and some noise). If these processes are also subject to an informational cascade, this may provide other possible loci for interactions not accounted for in the simulations.
The simulations leave out other factors that influence picture naming RT (e.g., AoA, word length, by-speaker variation).	If these factors are orthogonal to the effects of interest, then this variation would simply be an additional source of noise in the response times and would not affect the presence/absence of an interaction between codability and frequency. If these factors account for some of the same variability as codability and/or frequency measures, they may be other possible contributors to interactions that are not accounted for in the simulations.

Table 2.6 (Continued)

The simulations simplify the estimation of codability and frequency effects on lexical selection and phonological encoding as linear equations. These equations are unlikely to capture the precise magnitude of the effects for all individuals, circumstances, and languages.	For the goals of the present analysis, capturing the exact magnitude of the slowdowns in lexical selection and phonological encoding processes caused by codability and frequency is less important than capturing the way they interact. In additional simulations, we found that the overall pattern of results (no interactions for serial simulations, under-additive interactions for dynamic ones) is robust to different estimations of lexical selection and phonological encoding times. This suggests that the precise magnitude of the estimated codability and frequency effects in the simulations does not in and of itself produce the presence or absence of an interaction.
The simulations do not account for processing limitations such as processing load, working memory capacity, retrieval errors, or confusion.	These sources of noise in response times may be more likely to influence naming when codability is low (i.e., when there are more active lexical representations). If this is the case, these factors should make the codability effect more pronounced and may slow other processes like phonological retrieval when there is low name agreement, potentially contributing to an *over-additive* interaction between lexical selection and phonological encoding processes. If such processes are at play in the real world and retrieval is otherwise serial, then we'd expect no interaction or an over-additive interaction. This is ruled out by the data. If these processes exist in the context of cascading architecture, then our simulations (which do not include these effects) would overestimate the under-additivity of the observed interaction in the dynamic simulation, resulting in a more pronounced fan shape. Retrieval errors that slow RTs may be less likely to occur for more frequent forms, which should make the frequency effect more pronounced. How this affects the interaction is dependent upon the locus of this frequency effect.

Table 2.6 (Continued)

The simulations assume clean and instantaneous transmission of information between lexical selection and phonological encoding processes. The simulations do not assume an activation threshold for lexical items that must be reached before activation spreads to phonological form.	Non-instantaneous information transfer is an additional source of noise in the response times that is unlikely to cause or inhibit an interaction between codability and frequency. An activation threshold for information spread may increase the length of lexical selection by requiring that additional processing must occur before phonological encoding can start. This lengthening may be more likely in cases of low codability when there are more candidate names receiving activation. Lengthening lexical selection in cases of low codability is unlikely to produce an interaction in a serial framework. In a dynamic framework, this lengthening may lead to an even greater under-additive interaction.
The simulations do not model the potential effect of phonological competition on the speed of phonological encoding (e.g., in a resource-limited parallel model) or the potential role of activation feedback from phonological forms to their lexical representations.	In the dynamic simulation, adding phonological competition could potentially accentuate an under-additive interaction between codability and frequency, as competition would result in increased slowing in cases of low codability (when the forms of many lexical items become activated) that may attenuate any potential processing boost the target has from form frequency. In the serial simulation, if effects of codability occur before and separate from phonological encoding, adding phonological competition would not produce an under-additive interaction between codability and frequency on its own. We would not expect phonological competition to attenuate the size of the frequency effect for low name agreement (leading to an under-additive interaction) unless phonological neighborhood densities are larger for names applied to referents with lower codability.

Table 2.6 (Continued)

The simulations do not account for an effect of frequency on lexical selection.	The simulation likely overestimates the size of the phonological frequency effect. If frequency influences lexical selection in addition to codability, that provides the opportunity for the two factors to interact at that level, potentially producing an interaction within a serial architecture that is unaccounted for in our simulation. In the dynamic simulation, distributing the frequency effect across the lexical selection and phonological encoding processes should not eliminate the under-additive interaction between codability and the phonological frequency effect, though it may make the interaction less pronounced if the phonological frequency effect is smaller.

2.5 General discussion

In the present study, we analyzed naming times in five-year-old children and adults to understand how information flows between levels of representation in both the developing and mature language production systems. We asked whether the informational cascades present in the adult production architecture are a fundamental property of the language system present early in life or whether they only emerge later with experience. We addressed these questions in three studies.

Study 1 assessed the influence of codability (name agreement) and frequency manipulations on picture naming response time by adults and five-year-old children. By investigating these two factors, which have been argued to influence the processes of lexical selection (Alario et al., 2004; Griffin, 2001; inter alia) and phonological encoding (see Griffin & Bock, 1998), respectively, we can assess how these two processes relate to each other. Replicating prior results, we found that naming RTs were affected by both manipulations in both populations (Study 1, Q1). We observed slower naming times for items with low codability and

for low frequency names (see also Bates et al., 2003; Butterfield & Butterfield, 1977; D'Amico et al., 2001; Jescheniak & Levelt, 1994; Johnson, 1992; Johnson & Clark, 1988; Lachman, 1973; Lachman et al., 1974; Lachman & Lachman, 1980; Oldfield & Wingfield, 1965; Paivio et al., 1989; inter alia). Fitting ex-Gaussian distributions to the adult and child RT data (Study 1, Q2), we found that the codability and frequency manipulations engendered different effects on the RT distributions, supporting the hypothesis that they influence different processes. The RT distribution effects of the manipulations were consistent with their interpretations of influencing the processes of lexical selection and phonological encoding, respectively. These effects were qualitatively similar in both age groups, suggesting that there are similar underlying production planning processes at play in both the developing and mature language systems.

Critically, we also observed significant under-additive interactions between codability and frequency effects in both the adults and the five-year-olds: The size of the frequency effect was reduced when codability was low. To our knowledge, such an interaction has not been previously reported for either population. In Study 2, we demonstrated that this interaction is reliable across experiments and languages, documenting the same pattern of effects in Bates et al.'s (2003) English, German, Spanish, and Mandarin Chinese naming data. This under-additive interaction is predicted by a cascading model of word planning in which activation spreads from activated lexical representations to their phonological forms even before a lexical item has been selected for articulation (more below).

In Study 3, we confirmed our hypothesis about the source of the interaction by simulating the codability effect on lexical selection and the frequency effect on phonological encoding in different word planning architectures. We conducted two simulations: one in which lexical selection and phonological encoding processes occur simultaneously (approximating a simple

informational cascade) and another in which they occur sequentially (approximating a discrete, serial planning process). The simulations showed that an under-additive interaction between codability and frequency measures arises naturally in the first simulation but not the second. While this cascading model is not the only possible explanation for this interaction, we believe that given the prior evidence for cascaded processing in adults it is the most parsimonious explanation (e.g., Costa et al., 2000; Cutting & Ferreira, 1999; Jescheniak & Schriefers, 1998; Morsella & Miozzo, 2002; Peterson & Savoy, 1998; Rapp & Goldrick, 2000; Starreveld & La Heij, 1995; among many others). Thus, the presence of the same interaction in five-year-olds (Study 1) provides evidence that cascaded processing is already robustly present by this age, as we would expect it to be if this cascade is an inherent consequence of an incremental language production system.

In the remainder of the General discussion, we discuss potential interpretations of the interaction between codability and frequency effects (including alternative interpretations that do not assume cascaded processing), what our findings suggest for the development of cascading activation in the language production system, and questions for future research.

2.5.1. Interpreting the codability and frequency interaction

The presence of an interaction between image codability and response frequency in picture naming is intriguing, as it requires that the underlying mechanisms affected by the two factors be able to influence each other. This is difficult to reconcile with strictly sequential models of language production because codability and frequency effects have generally been assumed to affect different processes within word planning (lexical selection and phonological

encoding, respectively). We propose that the observed interaction reflects an incremental informational cascade between lexical selection and phonological encoding processes.

In a cascading framework, activation spreads to the phonological forms of the lexical candidates activated during selection. When codability is low and lexical selection time is lengthened, the phonological forms of the name alternatives (including the name ultimately selected for articulation) have more time to receive a cascading activation boost before selection of a lexical candidate. Given that the frequency of a word form influences the time it takes to access the phonological form, the size of the frequency effect should be diminished when the form of the selected name is already partially activated before phonological encoding, compared to cases when the word form receives less cascading activation and starts phonological encoding from closer to its base activation level. Viewed from a different perspective, when lexical selection is slowed due to low codability, word frequency's RT influence on phonological encoding will become less pronounced, as relatively faster or slower access of phonological forms will be in part obscured by the slowing at the lexical selection level. Thus, a cascading activation architecture predicts the under-additive interactions we observed.[9]

A strict serial architecture, on the other hand, does not predict that the factors influencing lexical selection and phonological encoding will interact. In such an architecture, activation spreads to the selected word's phonological form only once competition between different

[9] As pointed out by a reviewer, under some assumptions, cascading activation may in fact predict an over-additive interaction. In particular, over-additivity could arise if (i) phonological encoding begins as soon as lexical candidates are active, (ii) phonemes linked to different lexical candidates compete with each other (but phonemes activated by the same candidate do not), (iii) phonological competition depends on frequency (high frequency competitors result in greater competition), and (iv) in cases of low name agreement, the frequency advantage for the produced name relative to alternatives is greater for high frequency words than low frequency words. Under these assumptions, competition during phonological encoding would have a minimal effect on high frequency, low codability items and the largest effect on low frequency, low codability items, thereby resulting in an over-additive interaction. Our data are inconsistent with this hypothesis.

lexical candidates is resolved during lexical selection. In this framework, codability should influence the speed of lexical selection, with longer selection times when there are more active lexical candidates (i.e., codability is low), but it should not influence the subsequent phonological encoding stage once a name candidate has already been selected for utterance. Thus, under a strict serial model of word planning, the two effects should be additive (see Sternberg, 1969, 2001 for discussion).

Nevertheless, the presence of such an interaction, interpreted in isolation, does not provide conclusive evidence for cascading activation. There are (at least) two other ways that such an interaction could arise. The first way weakens the central assumption of the serial activation hypothesis by allowing multiple lexical nodes to be selected but only when multiple names could apply to an image (e.g., Levelt et al., 1999), as is the case for low codability items. In this scenario, the phonological forms for multiple candidate names may be activated for low codability items, which could lead to increased competition between forms at the phonological level. What effect this would have on the frequency effect would depend on the frequency distribution of the alternative labels and the degree to which frequency plays a role in resolving the competition at this lower level. If competition decreases the frequency effect (by introducing other factors that limit phonological encoding and produce overall slowing), this model might result in an under-additive interaction and thus capture the present data pattern.

Second, the interaction could result from an interplay of codability and frequency effects during lexical selection. As previously mentioned, there is evidence that frequency may influence lexical selection in addition to phonological encoding (e.g., Finocchiaro & Caramazza, 2006; Jescheniak & Levelt, 1994; Johnson et al., 1996; Kittredge et al., 2008; Strijkers et al., 2010). For example, a frequency effect at the lexical selection level may enhance the base

activation of more frequent lexical candidates (Strijkers et al., 2010), which could allow for faster activation and competition resolution in favor of these frequent candidates. Such an effect has the potential to speed up selection (and consequently production) of high frequency lexical representations, particularly in cases of high codability when there is little competition from alternative lexical candidates. Highly codable items with high frequency names should thus receive frequency-related activation boosts during both the lexical selection and phonological encoding processes, speeding their RTs compared to their low frequency counterparts. It is possible that a lexically-based frequency effect would have a smaller effect when codability is low and there is greater competition from alternative lexical candidates, leading to an under-additive interaction. To support this alternative account over a cascading hypothesis, one must determine whether the size of a lexical selection-based frequency effect is large enough to produce an interaction effect on its own (frequency effects on lexical selection have been argued to be smaller and less reliable than the frequency effect on phonological encoding; Griffin & Bock, 1998) and whether the size of such an effect is indeed diminished when there are multiple lexical candidates.

In sum, while an informational cascade between lexical selection and phonological encoding is one possible explanation for the interaction we observed, this explanation is not the only way such an interaction could arise (particularly given the complexity of attributing effects of manipulated variables to specific levels of processing). Nevertheless, this explanation relies on the fewest unproven auxiliary assumptions and is consistent with the other evidence for cascading processing in adults (e.g., Costa et al., 2000; Cutting & Ferreira, 1999; Jescheniak & Schriefers, 1998; Morsella & Miozzo, 2002; Peterson & Savoy, 1998; Rapp & Goldrick, 2000; Starreveld & La Heij, 1995).

2.5.2. Cascading activation in the developing language production system

Our findings in adults provide convergent evidence for a theory of mature language production that is well supported by prior studies of speech errors and interference paradigms — a theory in which there is cascading activation between lexical selection and phonological encoding. Our findings for children demonstrate that this same theory can explain the pattern of picture naming times in five-year-olds. Specifically, in young children, as in adults, there is an under-additive interaction between the codability and frequency effects, which is most readily explained by the continuous spread of activation from lexical selection to phonological encoding. This is important because prior studies directly addressing this question have been limited to children seven years of age and older (e.g., Jescheniak et al., 2006; Poarch & van Hell, 2012; Sylvia, 2017). The fact that cascaded processing is present this early in development suggests that it is a fundamental property of the language system as opposed to a property that emerges only with experience or adult-like cognitive processing.

Critically, however, the present study does not resolve the question of when and how cascaded processing develops in young children; it merely places new constraints on the answer. Children begin mapping words to meanings as young as six months of age (Bergelson & Aslin, 2017; Bergelson & Swingley, 2012; Bergelson & Swingley, 2013; Bergelson & Swingley, 2015; Tincoff & Juscyzk, 1999; Tincoff & Juscyzk, 2011). Their interpretation of words appears to improve substantially by 13–14 months (Bergelson & Swingley, 2012; Bergelson & Swingley, 2013; Bergelson & Swingley, 2015). In the second year of life, there are further improvements in the speed of lexical processing (Fernald et al., 1998; Fernald et al., 2006; Zangl et al., 2005) and changes in the robustness of phonological representations (Mills et al., 2004; Werker et al., 2002). Thus, comprehension studies suggest that the most substantial changes in lexical

representations may be occurring between roughly 12 and 24 months of age. Does cascading processing in lexical production emerge with these early changes in language comprehension? Or does it rely on later changes in the lexical network, like those associated with lexical prediction (Mani & Huettig, 2012)? The present study provides strong motivation for pursuing these questions in even younger children.

In addition, our findings bear on the question of how the effects of cascading activation change during the middle childhood and adolescence. As we noted in the Introduction, Jescheniak et al. (2006) observed preliminary evidence for a developmental change in cascading activation. They found evidence of semantically-mediated phonological activation (a predicted consequence of a cascaded processing architecture) in seven and eight-year-old children but not in nine and ten-year-old children or adults, which could suggest that the informational cascade between lexical representations and phonological forms is more robust in younger children (Jescheniak et al., 2006). Similarly, visual inspection of the interaction effect plots from our naming experiment (Figures 2.3 and 2.4) suggests that the interaction between codability and frequency may be more pronounced for five-year-olds than for adults. This observation is supported in the categorical analysis by the relative magnitudes of the beta coefficients for the interaction in the adult (β = -0.03) and child (β = -0.06) data. Moreover, in an exploratory categorical analysis, there was a three-way interaction between codability, frequency, and population (β = 0.02, $t(160) = 3.66$, $p < 0.001$) (though this contrast is not obviously reliable using continuous measures; see Supplementary Materials). If we assume that the interaction between codability and frequency reflects an informational cascade, this pattern would also suggest a developmental change in the strength of these effects.

A stronger cascade earlier in development could result from weaker inhibition of cascaded activation in younger children, resulting from a general deficiency in inhibitory capacity compared to older children and adults (see discussion in the *§2.1. Introduction*). Under this hypothesis, evidence of a cascade should become less pronounced as speakers become better able to suppress the activation of non-target forms. In fact, the child participants in Study 1 produced a greater proportion of false starts and renaming errors in the low codability conditions (1.6% of responses) than adults (0.6%), which may reflect such weaker inhibition of alternative name candidates (see Supplementary Materials for error distributions). Alternatively, a stronger cascade in younger children could result from slower overall processing, allowing more activation to cascade to non-target forms (see *§2.1. Introduction*); as production processes become more efficient and faster over development, non-target forms may receive less activation from the informational cascade, making their activation more difficult to detect.

Further research is necessary to identify the developmental trajectory of cascaded processing, including assessing whether it is possible to find evidence for informational cascades in even younger children and investigating possible changes in the strength of the cascade over time. Tracking the size of the codability and frequency interaction in naming RTs across different age groups has the potential to inform whether the strength of informational cascades changes over the course of development. One of the advantages of the picture naming paradigm we used in our study is that it is simpler than the priming and interference tasks typically used to look for evidence of cascaded processing in adults. The straightforward nature of the picture naming task makes it easy to run with child populations, providing a potential avenue to explore when cascaded processing first arises.

In addition to investigating interaction effects between codability and frequency, researchers could use this paradigm to look for other signatures of phonological form activation of multiple lexical candidates. For example, researchers could investigate children's production of false start and renaming errors, which may be more likely to occur when activation cascades to candidates other than the ultimately produced target name. Future work could manipulate additional properties of the images to be named that may influence the level of competition experienced in cases of an informational cascade, such as the relative frequencies and/or the phonological neighborhood densities of their candidate names. Tracking speech errors longitudinally may provide insight into the strength of the informational cascade and/or inhibition of such cascaded activation over development.

Nevertheless, the interaction we observed and the presence of speech errors, while consistent with an informational cascade, do not prove the existence thereof. A remaining question for future research is whether it is possible to obtain more direct evidence for informational cascades in children under seven years of age — for instance, by finding evidence of semantically-mediated phonological activation. Ideally, future research will additionally explore a wider range of data sets to assess not only when this property arises but also whether there are cross-linguistic differences in development or in the end state of the cascaded word planning architecture.

2.5. Conclusion

In the present study, we address the question of how early cascading activation emerges in the language production system. While prior work has shown evidence of informational cascades in children seven years of age and older (e.g., Jescheniak et al., 2006; Poarch & van

Hell, 2012; Sylvia, 2017), our study provides preliminary evidence that a cascading architecture is already in place by at least by five years of age. In a picture naming experiment, we observed qualitatively similar response time effects in both adults and five-year-old children, suggesting that similar underlying word planning processes are at play in both the developing and adult language systems. Critically, we observed under-additive interactions between image codability and name frequency effects in both populations. This interaction generalizes across experiments and languages and arises naturally from a word planning architecture in which activation cascades between lexical and phonological representations. Our study thus provides evidence for early cascaded processing, supporting the hypothesis that it is a fundamental property of the language system, rather than a capability that emerges gradually with experience.

Chapter 3

[Paper 2]

Evidence for top-down constraints and form-based prediction in four and five-year-olds' lexical processing

Margaret Kandel, Nan Li, & Jesse Snedeker

3.1. Introduction

Modern models of cognition involve multiple levels of representation, with both top-down and bottom-up mechanisms that pass information across these levels. Such interactivity is thought to be present in many cognitive processes, including visual perception (e.g., Bullier, 2001; Hochstein & Ahissar, 2002; Kafaligonul et al., 2015), olfactory perception (e.g., Andersson et al., 2018), gustatory perception (e.g., Kobayashi et al., 2004), somatosensory perception (e.g., Haegens et al., 2011), object recognition (e.g., Fenske et al., 2006; Panichello et al., 2012; Wyatte et al., 2014), face perception (e.g., Rossion et al., 2003; Sorger et al., 2007), and language (e.g., Dell, 1986; Dell & O'Seaghdha, 1991, 1992; Dell et al., 1997; Harley, 1993 for production; e.g., McClelland & Elman, 1986; Tanenhaus et al., 1995 for comprehension). An important question for cognitive scientists is how and when these complex interactions develop. Are the mechanisms for informational flow in both directions in place from early in life or do the necessary pathways develop later, perhaps with experience or as the brain matures?

Language serves as a particularly rich domain for investigating questions about interactive processing, since it involves multiple distinct levels of processing that can interact, each with qualitatively different types of representations. For example, language comprehension involves perceptual processing (processing the auditory or visual linguistic input), phonological processing (segmenting the linguistic input into discrete units), prosodic processing (identifying the prosodic cues in the input), lexical processing (mapping the processed input onto stored representations and retrieving these representations from the mental lexicon), syntactic processing (using prosodic and lexical information to identify how words are grouped in phrases and linked together), semantic processing (interpreting the identified lexical and syntactic representations), and pragmatic processing (integrating semantic information with the discourse or broader context). All of these processing levels serve a different function and involve different representations. Information flows through these levels incrementally and in a cascaded fashion, meaning that representations are activated at each level even as the input is still unfolding, and information passes between levels of processing before processing at the preceding level is complete (e.g., Allopenna et al., 1998; Huang & Snedeker, 2011; Marslen-Wilson, 1987; Ferreira & Clifton, 1986; Garnsey et al., 1997; Yee & Sedivy, 2006; inter alia). Within this framework, the flow of information from lower levels of representation (those closer to articulatory or perceptual processes) to higher levels (those representing complex meanings) is referred to as bottom-up processing; the flow of information in the opposite direction is referred to as top-down processing.

Although language comprehension can be viewed as a largely bottom-up process, in which speakers start with perceptual representations of the input and build a representation of the conceptual message intended by the speaker, there is evidence that processing at many levels is

influenced not only by preceding levels of processing (bottom-up influences) but also by information at subsequent levels (top-down influences). This interactivity allows information about likely sentence meanings, syntactic context, and the speaker's intentions to influence the processing of words that appear in the linguistic input. In particular, adult language users are able to use top-down contextual information to rule out incongruent lexical candidates during word recognition (e.g., Dahan et al., 2000; Dahan & Tanenhaus, 2004; Magnusson et al., 2008) and even to pre-activate, or predict, upcoming word representations before they are encountered in the input (see Ito, 2024; Kuperberg & Jaeger, 2016; Ryskin & Nieuwland, 2023 for review). This interactive processing makes language comprehension more efficient and robust to degraded input.

The present study investigates when this interactivity develops by exploring the degree to which top-down information influences the lexical processing of four and five-year-old children. We focus on two measures of top-down influence: (i) the use of contextual information to guide word recognition and (ii) phonological form prediction. We compare these phenomena in children and adults. Understanding the early stages of the language comprehension system and the ways that they resemble or differ from mature language processing can provide insight into the system's architecture and is crucial for constructing theories of comprehension (and cognition more generally) that are compatible with all life stages.

The remainder of this Introduction is divided into two parts. First, we discuss the processes that underlie spoken word recognition, how the availability of top-down contextual cues influences these processes in adults, and why such interactive processing may be challenging for young children. Second, we explain how phonological form prediction differs from the type of prediction assessed in most prior eye-tracking studies with children (i.e.,

referential prediction), and we discuss the available evidence for phonological form prediction in adults and children.

3.1.1. Spoken word recognition in auditory language comprehension

During spoken word recognition, listeners process bottom-up acoustic-phonetic input incrementally, rather than waiting until they have heard a complete word before attempting to map it to a stored representation (e.g., Allopenna et al., 1998; McClelland & Elman, 1986; Norris 1994). As the acoustic-phonetic input unfolds, it activates candidate words from the listener's mental lexicon that match this input. In the absence of any top-down expectations, this leads listeners to initially consider words starting with the same sound sequence as the spoken word (i.e., the word's phonemic cohort) as potential candidates during word recognition, eventually narrowing down to a single word candidate once the input disambiguates between options (though note that candidates are not strictly restricted to those with matching onsets; see Allopenna et al., 1998 for rhyme effects).

This incremental bottom-up processing is evident in visual world eye-tracking experiments in what is known as the *phonemic cohort effect*. In the typical visual world paradigm, participants view a display, and their eye movements are recorded as they listen to an utterance; these eye movements provide insight into language comprehension processes, as individuals systematically look to referents or associates of the words they hear (e.g., Cooper, 1974; Tanenhaus et al., 1995; see Huettig et al., 2011 for review). In visual world experiments, when a listener hears a spoken target word (e.g., *beaker*) that shares onset phonemes with the name of one of the images on the screen (e.g., *beetle*), they initially fixate on this cohort competitor more than phonologically-unrelated distractors (e.g., *carriage*), only looking away

once they've received the acoustic-phonetic input required to disambiguate the competitor from the target (Allopenna et al., 1998). This gaze pattern indicates that listeners process the unfolding target word incrementally, initially activating multiple candidates that match the acoustic-phonetic sequence they encounter. The phonemic cohort effect has been replicated many times with both adults (e.g., Allopenna et al., 1998; Dahan & Gaskell, 2007; Dahan et al., 2001; Farris-Trimble & McMurray, 2013; Magnusson et al., 1999; inter alia) and young children (e.g., Desroches et al., 2006; Kandel & Snedeker, 2024; Rigler et al., 2015; Sekerina & Brooks, 2007; Weighall et al., 2017; inter alia), indicating that incremental bottom-up processing begins early. In fact, incremental bottom-up processing is present by the second year of life (Fernald et al., 2001; Swingley, 2009; Swingley et al., 1999). The onset time of competition effects is similar in preschoolers and adults, but the effects often last longer in children (Sekerina & Brooks, 2007).

3.1.1.1. Top-down influence in adult word recognition

There is evidence that adult listeners can use top-down information about a word's context to guide the incremental bottom-up processing of acoustic-phonetic input during word recognition — this information allows them to rule out incongruent lexical candidates, even if they initially match the input. In visual world experiments, this integration of top-down information leads to a modulation of the phonemic cohort effect based on the target word's context. For instance, Dahan and Tanenhaus (2004) observed an effect of semantic context on cohort competitor activation. They manipulated whether a target word (e.g., *bok* [goat]) was preceded by a sentence context containing a verb that was semantically constraining (e.g., *Nog nooit* ***klom*** *een bok zo hoog* [Never before **climbed** a goat so high]) or neutral (e.g., *Nog nooit* ***is*** *een bok zo hoog geklommen* [Never before **has** a goat climbed so high]). They then investigated

looks to a semantically-unrelated cohort competitor (e.g., *bot* [bone]) as participants heard the target word. Dahan and Tanenhaus (2004) observed a typical phonemic cohort effect in the neutral verb condition, but not in the constraining verb condition, suggesting that listeners were able to use the target word's semantic context to rule out semantically-incongruent word candidates (like *bone*). Dahan et al. (2000) additionally found evidence that listeners can use syntactic context to guide word recognition: When a target word was preceded by a gender-marked definite article, listeners did not look to gender-inconsistent cohort competitors more than phonologically unrelated distractors, suggesting that they used the grammatical gender cues provided by the definite article to constrain word candidates and avoid competition. Magnusson et al. (2008) and Strand et al. (2018) further found that listeners can use syntactic category expectations to avoid cohort competition from wrong-category competitors (though cf. Gaston, 2020). Note that these effects could in theory result from facilitation of congruent word candidates (boosting their activation, allowing for easier access) or inhibition of incongruent candidates (preventing their activation, thereby restricting the candidate set and allowing for easier access of congruent candidates). For convenience, we refer to these effects as avoiding competition from incongruent candidates; we do not make any commitments to the activation mechanism of this effect in the present paper.

Recent studies have found that cohort effects at target word onset in constraining contexts are absent even without the target image on the screen (Ito et al., 2018; Li et al., 2022). This is important because when the target is depicted on the screen, the absence of a phonemic cohort effect in contextually constraining conditions could result in part from the presence of anticipatory target looks due to referential prediction (see section *1.2 Using top-down information to predict phonological form* for details). For instance, in Dahan and Tanenhaus

(2004), the semantic association between the constraining verb (e.g., *climb*) and the target (e.g., *goat*) may have allowed participants to identify the target as the most likely image to be referred to later in the sentence, leading them to launch or plan looks to the target image even before hearing the target word input (a verb-based anticipatory looking effect; e.g., Altmann & Kamide, 1999), thereby reducing the amount of cohort competitor looks at target word onset. Indeed, Dahan and Tanenhaus (2004) observed increased fixations to the target image much earlier than expected (~100 ms after target word onset) given saccade planning and execution times in response to linguistic stimuli in other experiments (~200 ms after target word onset; e.g., Allopenna et al., 1998; Cooper, 1974). Furthermore, whether the visual world display includes a representation of the target word appears to have an effect in studies of syntactic constraint (see Strand et al., 2018 vs. Gaston, 2020). Thus, studies with a cohort member but no target on the screen provide a stronger test of the interaction between bottom-up and top-down sources of information during lexical processing. The present experiment goes beyond the prior adult studies of this kind (Ito et al., 2018; Li et al., 2022) by including both constraining and neutral sentences, allowing us to confirm cohort activation in the absence of strong top-down constraints.

3.1.1.2. Top-down influence in children's bottom-up language processing

Prior work with adults paints a compelling picture of how top-down information can guide processing of the bottom-up input during word recognition. In particular, words whose meanings are not congruent with the context do not appear to be activated as potential matches to the bottom-up input, even if the word matches the initial sequence of phonemes in the input. Interactive processing of this kind requires comprehenders to rapidly construct higher-level

representations of sentence meaning and quickly pass that information down so that it can inform word recognition. This process may present a challenge for young children, who have less language experience, poorer executive functioning (Best & Miller, 2010; Mazuka et al., 2009), smaller working memory capacity (Chi, 1978; Dempster, 1981; Schneider & Bjorklund, 1998), reduced multitasking ability (Yang et al., 2017), and slower processing speed (Hale, 1990; Kail, 1991; Kail & Salthouse, 1994).

In line with this hypothesis, prior research suggests that top-down processing is challenging for young children in some domains. For instance, visual world studies of syntactic ambiguity resolution find that four to six year-old children show reduced use of top-down information to guide syntactic parsing, relying instead on bottom-up cues, such as intonation and lexical information (e.g., Kidd et al., 2011; Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008; Trueswell et al., 1999; Yacovone et al., 2021). In addition, school-aged children (7–12 years) are less likely than adults to use top-down information to disambiguate homophones (Khanna & Boland, 2010; but cf. Hahn et al., 2015) or facilitate lexical access while reading (e.g., Joseph et al., 2008; Tiffin-Richards & Schroeder, 2020). These data patterns could indicate that top-down and bottom-up language comprehension pathways develop separately or that information only moves through these pathways rapidly enough to interact once an individual has sufficient experience, memory capacity, or processing efficiency. In fact, differences in language experience, memory, and processing speed have been found to influence contextual effects in adults (e.g., Huettig & Janse, 2016; Ito et al., 2018; Li & Qu, 2024).

However, as discussed in the next section, there is evidence from recent electroencephalography (EEG) experiments that children (5–10 years) do use top-down cues to facilitate lexical processing in naturalistic listening tasks (e.g., Levari & Snedeker, 2024), which

could suggest that the pathways required for interactive activation flow are in place, even if they are not used in all tasks or for all processes. Thus, it remains unclear whether young children have the capability to use top-down information to guide word recognition the way that adult listeners do.

3.1.2. Using top-down information to predict phonological form

As comprehenders read, listen, or watch a sentence unfold, they make predictions about things that they have not yet encountered. This ability arises early in life, by the age of two years (e.g., Borovsky et al., 2012; Borovsky et al., 2014; Lukyanenko & Fisher, 2016; Mani & Huettig, 2012; Nation et al., 2003; Özge et al., 2019; Sommerfeld et al., 2023; inter alia; see e.g., Altmann & Kamide, 1999 for adults). For instance, when two year-olds are presented with a display containing an image of a bird and a cake, they will look more to the cake than the bird after hearing a verb whose meaning is semantically constraining towards the cake (e.g., *Der Junge **ißt** den großen Kuchen* [The boy **eats** the big cake]) but not when the verb is neutral (e.g., *Der Junge **sieht** den großen Kuchen* [The boy **sees** the big cake]) (Mani & Huettig, 2012). However, in such anticipatory looking tasks, it is not obvious that what is predicted is lexical in nature. While looks could be guided by the lexical expectations, they could also be guided by expectations about the semantic space and the events that are likely to occur. Listeners may look at an image of a cake after hearing the verb *eat* because cakes are edible, and listeners think they are likely to be eaten, as opposed to looking at the cake because they are predicting that they will hear the word *cake* in the input. In fact, contextually-mediated looks to referent images can arise even in contexts in which it is unlikely that the referent will appear as a word in the sentence, such as in wh-questions (e.g., looks to an image of a fork after hearing the word *cake* in the

sentence, *What was Emily eating the cake with?*; Atkinson et al., 2018; Seidl et al., 2003). Thus, while anticipatory target looks reflect a kind of prediction (which we refer to as *referential prediction*), they do not reveal whether listeners make top-down lexical predictions specifically and whether these predictions extend to the level of a word's phonological form features. This is the question we are addressing in the present study.

3.1.2.1. Phonological form prediction in adults

There is evidence that adults can use top-down contextual information to pre-activate, or predict, specific word representations before they are encountered in the input (see Ito, 2024; Kuperberg & Jaeger, 2016; Ryskin & Nieuwland, 2023 for review). Much of the evidence for lexical prediction is derived from EEG experiments, which measure event-related potentials (ERPs), or changes in electrical voltage on the scalp produced by the electrical activity of the brain that are time-locked to the onset of linguistic stimuli. These experiments show reduced N400 responses (negative-going changes in the ERP waveform peaking at approximately 400 ms after stimulus onset; see Kutas & Federmeier, 2011 for review) to words that appear in supportive contexts (e.g., Bentin et al., 1985; Federmeier & Kutas, 1999; Ganis et al., 1996; Holcomb, 1988; Kutas, 1993; Kutas & Hilyard, 1984; Rabs et al., 2022; Van Petten, 1993; Wlotko & Federmeier, 2012; inter alia), suggesting that adults are able to use these contexts to predict upcoming word information.

While these ERP effects are consistent with lexical prediction, they can be explained by other mechanisms, such as rapid semantic integration (e.g., Brown & Hagoort, 1993). More convincing evidence for lexical prediction comes from studies that directly test whether adults

predict the form of an upcoming word (see Ito, 2024).[1] One source of evidence comes from studies that test ERP responses to words (or nonwords) that are phonologically similar to predictable target words (e.g., DeLong et al., 2019, 2020; Ito et al., 2016; Kim & Lai, 2012; Lazlo & Federmeier, 2009; Liu et al., 2006; Yacovone et al., 2024; inter alia). Studies taking this approach have found reduced N400s in constraining contexts (e.g., *She measured the flour so she could bake a*…) to words or nonwords that are phonologically similar to a target word (e.g., *ceke* for target *cake*) relative to unrelated words or nonwords (e.g., *tont*; Kim & Lai, 2012). These results suggest that the phonological features of the target had been pre-activated by the context, facilitating processing of the form-similar strings. Note that while some studies have found this effect to be dependent upon having a sufficiently slow presentation rate (Ito et al., 2016), there is evidence that form-based prediction also occurs in naturalistic settings with spoken language (Yacovone et al., 2024). Additional EEG evidence for form-based prediction comes from studies testing for pre-activation of word-form features prior to word onset using representational similarity analysis (e.g., Wang et al., 2024).

The question of phonological prediction in language comprehension has also been addressed using visual world eye-tracking studies (e.g., Ito & Husband, 2017; Ito et al., 2018; Kukona, 2020; Li et al., 2022; inter alia; see Ito, 2024 for review). These studies look for early phonemic cohort effects arising prior to articulation of the target word, referred to as *predictive phonological competitor effects* by Ito (2024). If listeners reliably look at phonemic cohort competitors more than unrelated distractors even before hearing the target word, that would suggest that they are able to predict and pre-activate the target word's phonological form before it appears in the input. For instance, Ito et al. (2018) presented participant with sentences

[1] In this discussion, we will not distinguish between prediction of orthographic and phonological word form.

constraining towards a predictable target word (e.g., *The tourists expected rain when the sun went behind the **cloud**, but the weather got better later*) along with a visual display that at times contained a phonological competitor of the target word (e.g., *clown*) (note that the target word was never depicted in the same display as the competitor). This visual display appeared 1000 ms prior to target word onset. Ito et al. (2018) found that native speakers looked predictively at the phonological competitor shortly after onset of the display but before target word onset, suggesting that the participants had pre-activated the target's phonological form (see, e.g., Li et al., 2022; Li & Qu, 2024 for similar effects). Although these predictive phonological competitor effects are inconsistent across studies (e.g., Ito & Husband, 2017; Ito & Sakai, 2021), a meta-analysis using 24 datasets from 20 different experiments revealed that the predictive phonological competitor effect was small but reliable and modulated by target word predictability (the effect was larger for more predictable words; Ito, 2024). Taken with the EEG results discussed above, it appears that adults have the capability to predict phonological form, particularly when given sufficient time and sufficiently constraining cues.

3.1.2.2. Phonological form prediction in young children

It is unclear whether we should expect to see evidence of phonological prediction in young children. Children's cognitive immaturity may mean that they do not have sufficient resources or processing speed to predict lexical information to the phonological level (which requires first pre-activating representations at the conceptual and lexical level), and their more limited language experience may make them less sensitive to the constraining cues available in the context. In addition, as discussed previously, young children have been observed to fail to incorporate available top-down contextual cues during other linguistic processes (e.g., Kidd et

al., 2011; Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008; Trueswell et al., 1999; Yacovone et al., 2021). If this is due to global limitations on top-down processing, we would expect children to rarely or never make phonological predictions.

Nevertheless, there is evidence that young children can make some kinds of predictions during language comprehension. As mentioned above, children engage in referential prediction by two years of age (e.g., Mani & Huettig, 2012), though the ability to make referential predictions based on phonological expectations appears to arise later (Gambi et al., 2018). By approximately three years of age, children show N400 effects that are sensitive to a word's semantic fit within its sentence context (e.g., Adamson-Harris, 2000; Johnson, 2011; Silva-Pereya, Klarman, et al., 2005; Silva-Pereyra, Rivera-Gaxiola, et al., 2005), which may reflect pre-activation of semantic features. As mentioned previously, however, such N400 effects could also be attributable to post-lexical semantic integration (e.g., Brown & Hagoort, 1993). Levari and Snedeker (2024) found using a naturalistic story listening paradigm that children five to ten years of age show N400 effects to word predictability, even when controlling for semantic associations with the context — thereby suggesting that in naturalistic settings, children in this age range make specific predictions about upcoming words based on top-down contextual information. However, it is not obvious from this study that these predictions extend to the level of phonological form. While there is evidence that five and six year-old children can make explicit predictions of high cloze words in a production task (Waite et al., under review), which requires retrieving the target word's phonology, it is not clear whether this occurs spontaneously or rapidly enough to result in prediction during online language comprehension. In short, while children are able to anticipate upcoming information during comprehension, there is no clear evidence in the literature that this ability extends to form-based prediction.

3.1.3. The present study

In the present study, we use visual world eye-tracking to look for evidence of top-down and bottom-up interactivity in the language comprehension of four and five-year-old children. This study will provide information about the age at which this interactivity develops. We divide this investigation into two questions:

Q1: Do four and five-year-old children use top-down contextual information to guide bottom-up processing during spoken word recognition, similar to adults?

Q2: Can four and five-year-old children additionally use contextual information to predict the phonological forms of upcoming target words?

We conducted the same experiment with children and adults so that we can directly compare their lexical processing. The design of our visual world task was inspired by that of Dahan and Tanenhaus (2024). In the task, participants heard sentences that were either highly constraining or neutral towards a target word. At times, the accompanying visual display contained a semantically-unrelated image whose name was a cohort competitor of the target. The target word was never depicted on the screen in order to avoid the potential of anticipatory target looks drawing looks away from the cohort competitor in the constraining sentences (e.g., Altmann & Kamide, 1999; Borovsky et al., 2012).

To answer Q1, we looked for a modulation of the phonemic cohort based on the target word's predictability. If we observe reduced or absent cohort effects in the constraining contexts compared to the neutral contexts, that would suggest that participants are able to use the contextual cues in constraining contexts to guide word processing, ruling out the cohort competitor as a potential match to the spoken input. To answer Q2, we looked for a predictive

phonological competitor effect shortly after the target words become predictable in the constraining sentences. If participants show an early phonemic cohort effect prior to target word onset, that would indicate that the target words' phonology had been pre-activated, providing evidence of phonological form prediction.

3.2. Methods

The methods and analyses for this study were preregistered on OSF (https://osf.io/uf3tp/). Additional analyses that were not preregistered are identified as post-hoc. Data, analysis code, and Supplementary Materials are available from https://osf.io/r6ctx/.

3.2.1. Participants

The primary sample included 56 child participants (M_{age} = 5.0 years, SD = 0.5, range = 4;1;3–5;9;29, 29 F, 27 M) and 56 adult participants (M_{age} = 20.6 years, SD = 4.1, range = 18–45, 39 F, 17 M) who were native, monolingual speakers of American English.[2] Participants reported having normal or corrected-to-normal hearing and vision. Child participants were recruited through the Harvard Laboratory for Developmental Studies database. Child participants were given a small toy for participating, and their parents were given $10. Informed written consent was received from the parent or guardian for their child's participation, and children provided written assent. Adult participants were recruited from undergraduate classes at Harvard University and received partial course credit for their participation. Adult participants provided informed written consent for their participation.

[2] Gender information was derived from parental or self report. A participant was considered to be a native, monolingual speaker of American English if they spoke American English fluently, started learning it before the age of six, and did not speak any other languages fluently that they had started learning before the age of six (fluency was determined via parental/self report).

An additional 21 child participants were run in the experiment but were omitted from the primary sample because they were outside of the 4–5 year age range (1), they did not pass our exclusion criteria (15; see *§3.2.4. Data processing & exclusion*), they fussed out (2), a technical error prevented them from completing the experiment (2), or their vision was not corrected to normal (1). An additional 15 adult participants were run in the experiment but were omitted from the primary sample because they did not meet the language criteria (13), they did not pass our exclusion criteria (1; see *§3.2.4. Data processing & exclusion*), or a technical error prevented them from completing the experiment (1).

To address Q2, we analyzed both the primary sample as well as an exploratory sample that included 11 additional child participants (total N = 67; M_{age} = 4.9 years, SD = 0.5, range = 4;0;12–5;9;29, 34 F, 33 M) and 1 additional adult participant (total N = 57; M_{age} = 20.6 years, SD = 4.1, range = 18–45, 40 F, 17 M). In the exploratory sample, we did not omit trials with incorrect image selections and did not omit participants for having high levels of image selection inaccuracy (see *§3.2.4. Data processing & exclusion*). This decision was made in order to avoid the potential of excluding participants and trials from our analysis in which phonological prediction was so strong that it increased pre-activation of the cohort competitor to such an extent that it was inaccurately selected in place of the target image (see *§3.4. Q2: Phonological prediction in constraining sentences* for details). The exploratory child sample included 9 participants who were omitted from the primary sample for high levels of inaccuracy and 2 participants who were omitted for trial loss. The exploratory adult sample included 1 participant who was omitted from the primary sample for high levels of inaccuracy. Note that the overall pattern of results of the Q1 analyses were the same in the primary sample and the exploratory sample.

3.2.2. Materials

In each trial of the experiment, participants were presented with a display of four images and an auditory sentence. The participants' task was to select the image in each trial that matched a word from the sentence (the task image). The images used in the experiment were colorized digital images from various sources, including the BCBL MultiPic databank (Duñabeitia et al., 2018) and the Snodgrass and Vanderwart "Like" Objects (Rossion & Pourtois, 2004). Sentences were digitally recorded by a native, monolingual speaker of American English and were normalized in Praat (Boersma, 2001) to a RMS (root mean square) of 68 dB.

The sentence stimuli for the experiment comprised 30 experimental sentence pairs and 18 filler sentences (see the Supplementary Materials). Each sentence included a task word that was represented by one of the images on the screen (the task image). Note that the task word was not the same as the target word in our analysis (more below). The purpose of the task word and image was to provide the participants with an image to select in order to complete the experimental task; this task prompted participants to look for images that matched the words they heard in the stimulus sentences. The filler sentences ensured that the position of the task word within the sentence varied across trials so that participants would look for image matches throughout the duration of the stimulus sentences.

The experimental sentences included a target word prior to the task word. The target word was not represented in the image display. Within each experimental sentence pair, we manipulated the context preceding the target word such that the target word was highly predictable in one of the two sentences (the constraining condition) but not the other (the neutral condition) (Table 3.1). In the constraining sentences (e.g., *The baby drank the* ***milk****…*), average target word cloze was 75% (SD = 15%, range = 50–100%). In the neutral sentences (e.g., *The*

*child took the **milk**…*), average target word cloze was 2% (SD = 3%, range = 0–10%) (see *§3.2.2.1. Sentence-level controls* for more detail about stimulus norming). The same number of syllables appeared before the target word in both the constraining and neutral conditions for each sentence pair (M = 10 syllables, SD = 3, range = 5–19). The experimental sentences were rotated across presentation lists such that only one sentence per pair (constraining or neutral) appeared in each list. Within each list, half of the sentences were in the constraining condition, and half were in the neutral condition.

The visual displays for the filler trials included the task image (corresponding to the task word) and three unrelated distractor images. The visual displays for the experimental trials included the task image, two unrelated distractors, and a competitor image (Figure 3.1). Each target word in the experiment was assigned a cohort competitor word with an onset overlap of one or more phonemes (e.g., target: *milk*, cohort competitor: *mitten*) (M overlap = 2.4 phonemes, SD = 0.7, range = 1–4). The set of competitor images consisted of depictions of these cohort competitor words. We manipulated whether the competitor images appeared with a sentence containing their corresponding target word (the cohort condition; e.g., competitor *mitten* with target *milk*) or with a different sentence containing an unrelated target word (the control condition; e.g., competitor *mitten* with target *leash*). Competitor images were pseudorandomly assigned to their control targets such that the depicted competitor word was both phonologically and semantically unrelated to the control target word. Within each presentation list, competitor images appeared only once and were evenly divided between the cohort and control conditions.

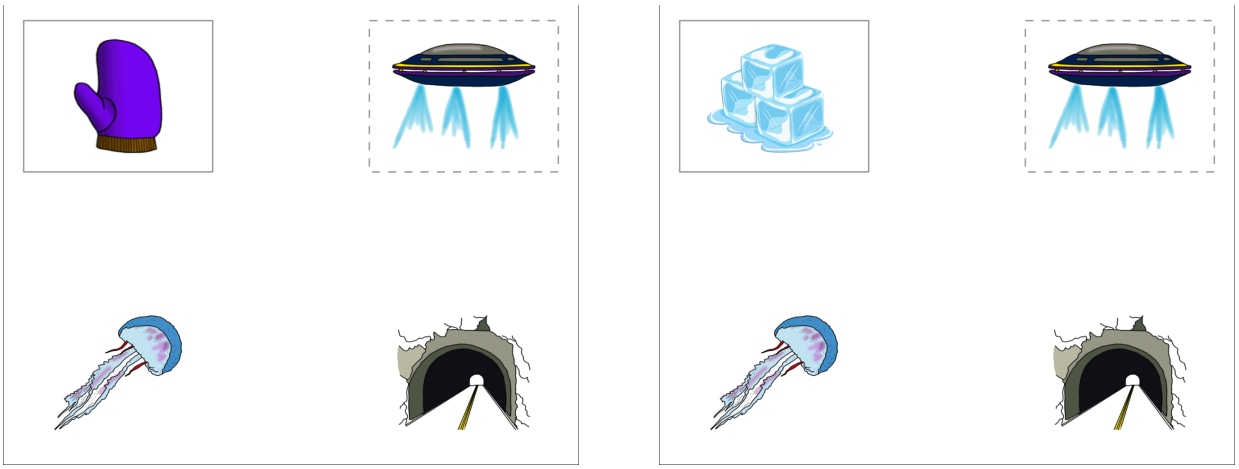
Images were displayed on a 17-inch Tobii T60 screen with a 1280 x 1024 pixel resolution. Images were printed with a size of 320 x 256 pixels (corresponding to 25% of screen width and height) and were positioned in each quadrant of the screen. Image centers were

aligned at 20% monitor width and 20% monitor height (top left), 80% monitor width and 20% monitor height (top right), 20% monitor width and 80% monitor height (bottom left), and 80% monitor width and 80% monitor height (bottom right). Image locations were fixed for each item (i.e., for a particular experimental sentence pair, the competitor, task, and distractor images would always appear in the same quadrants). In the experimental trials, the competitor image always appeared in one of the two quadrants horizontally opposite the side of the screen containing the task image. Image locations were pseudorandomly assigned to balance how often the task images and competitor images appeared in each quadrant.

Table 3.1: *Example stimulus sentences.* In each sentence, the target word is bolded, and the task word is underlined. Each target was paired with a cohort and a control competitor. Control competitors were cohort competitors for other targets (e.g., the cohort competitor *mitten* with target word *milk* was the control competitor for the target word *leash*).

Sentence condition	Example sentence	Cohort competitor	Control competitor
Constraining	The baby drank the **milk** that his mother prepared for him and then played with his toy spaceship.	mitten	ice
Neutral	The child took the **milk** that his mother prepared for him and then played with his toy spaceship.	mitten	ice
Constraining	The dog walker put the dog on a **leash** and walked down the street to the corn maze.	leaf	mitten
Neutral	The person on the sidewalk held a **leash** and walked down the street to the corn maze.	leaf	mitten

Figure 3.1: *Example cohort trial (left) and control trial (right).* In each display, the competitor image is marked with a solid rectangle, and the task image is marked with a dashed rectangle. This figure includes images from the BCBL MultiPic databank (Duñabeitia et al., 2018), the Snodgrass and Vanderwart "Like" Objects (Rossion & Pourtois, 2004), and Pixabay (https://pixabay.com/).



3.2.2.1. Sentence-level controls

We conducted three norming tasks to control for the predictability, congruency, and referential likelihood of the target, cohort competitor, control competitor, task, and distractor words in our experimental sentences (see Supplementary Materials for details). The results of these tasks are summarized in Table 3.2. In the constraining sentences, the target words were highly predictable and served as congruent sentence continuations. In the neutral sentences, none of the words were highly predictable, and the target, cohort competitor, and control competitor words were all congruent sentence continuations (though the target was the most congruent). Importantly, the cohort and control competitors had similar congruency and likelihood ratings to each other in both the constraining and neutral sentence conditions (p's > 0.05 in Welch two sample t-tests), suggesting that any observed differences between competitor image conditions

(cohort, control) within the sentence conditions would be due to the difference in their phonological relatedness to the target. The cohort competitor, control competitor, task, and distractor words were considered unlikely to be referred to later on in the sentences at the point at which participants encounter the target word, suggesting that the corresponding images would not draw anticipatory looks during our analysis window.

We used Latent Semantic Analysis (LSA; http://wordvec.colorado.edu; Landauer et al., 1998; Wolfe & Goldman, 2003) to additionally ensure that the associations between these critical words were low ($|LSA| \leq 0.15$). Target words were not semantically related to the cohort competitor ($M = 0.06$, $SD = 0.04$), control competitor ($M = 0.07$, $SD = 0.06$), task ($M = 0.06$, $SD = 0.05$), or distractor ($M = 0.05$, $SD = 0.05$) words. Cohort and control competitor words were also not semantically related to each other ($M = 0.06$, $SD = 0.04$) or to the task words (cohort $M = 0.05$, $SD = 0.05$; control $M = 0.05$, $SD = 0.05$) or distractor words (cohort $M = 0.05$, $SD = 0.05$; control $= 0.05$, $SD = 0.05$).

Table 3.2: *Results of the stimulus norming tasks showing the cloze values, congruency, and likelihood ratings for the target, competitor, task, and distractor words (as applicable) for the constraining and neutral sentence conditions.* The summary statistics in the table reflect the means of by-item values.

	Constraining Sentences			Neutral Sentences		
Word	Cloze (%)	Congruency (scale 1–5)	Likelihood (scale 1–5)	Cloze (%)	Congruency (scale 1–5)	Likelihood (scale 1–5)
target	75 (SD 15)	4.94 (SD 0.10)	N/A	2 (SD 3)	4.06 (SD 0.72)	N/A
cohort competitor	0	1.43 (SD 0.47)	1.51 (SD 0.47)	0	3.34 (SD 0.83)	2.68 (SD 0.66)
control competitor	0	1.54 (SD 0.63)	1.56 (SD 0.57)	0 (SD 1)	2.95 (SD 0.81)	2.39 (SD 0.71)
task	0	N/A	1.69 (SD 0.59)	0	N/A	2.08 (SD 0.51)
distractors	0	N/A	1.42 (SD 0.31)	0 (SD 1)	N/A	1.91 (SD 0.48)

3.2.2.2. Audio splicing

The audio stimuli for the experimental sentences were spliced so that the acoustic cues leading up to the target and task words would be comparable across sentence conditions. There were two versions of the audio for each constraining and neutral sentence. In one version, sentence material was spliced into a base recording from another recording of the same sentence (e.g., constraining base + constraining splice). In the other version, sentence material was spliced from a sentence of the other condition (e.g., constraining base + neutral splice). The splice recordings for each sentence were different from the base recordings (e.g., constraining base ≠ constraining splice). We spliced into the base recordings the parts of the sentence that were the same in the constraining and neutral conditions leading up to and following the target word; this always consisted of the target word itself and typically also included words leading up to the

target. For sentences that differed across conditions between the target and task words, we spliced the parts of the sentence that were the same in both conditions before and after the target word as well as before the task word. The audio stimuli for the filler sentences were spliced with an alternative recording of the same sentence prior to the task word so that both experimental and filler sentences would contain splicing.

3.2.3. Procedure

Participants were seated at a comfortable distance in front of a Tobii T60 infrared eye-tracker with a MagicTouch touch screen extension. At the start of the experiment session, participants completed a five-point calibration procedure using Tobii Studio software (v2.0.4).[3] The experiment was run in E-Prime 2.0 (Psychology Software Tools, Inc.) from a Dell Latitude E6410 host laptop. Participants were told that in each trial of the experiment, they would see four images and hear a sentence including the name of one of the images; they were instructed to select this image using the touch screen.

The experimenter checked the Tobii track status before initiating each trial. Once the trial was initiated, the image display appeared on screen. After 2000 ms, a small blue circle (30 x 30 pixels) appeared in the center of the screen, and participants heard an audio instruction telling them to fixate on it ("Look at the circle"). The fixation circle remained on screen for 2000 ms. After the circle disappeared, the auditory sentence played over an external speaker positioned

[3] Due to a technical error with Tobii Studio, we were not able to initiate the calibration sequence for two child participants (one of these participants was excluded from the primary sample but included in the exploratory sample). We used the participants' mean task image looks from task word onset to trial offset as a metric to evaluate whether the lack of calibration affected the quality of their eye-tracking data. The uncalibrated participants' mean task image looks were within one standard deviation of the means of the calibrated participants, suggesting that the eye-tracker detected their gaze with sufficient accuracy for analysis.

behind the Tobii. After the end of the sentence, an image of a magnifying glass appeared in the center of the screen. At this point, participants made their image selection using the touch screen.

Participants were randomly assigned to one of eight presentation lists of 48 trials (30 experimental, 18 filler). These presentation lists ensured that each competitor image appeared in each of the four conditions created by our 2 × 2 manipulation of sentence condition (neutral, constraining) and the competitor's relation to the target word (control, cohort) as well as with both audio versions for each sentence (see *§3.2.2.2. Audio splicing*). Trial order was randomized for each participant.

Three practice trials were presented at the beginning of the experiment. During the first practice trial, the experimenter explained the different trial stages that the participants would encounter as they completed the experiment. The next two practice trials followed the same format as the experimental block. All participants saw the same three practice trials in the same order. As participants completed the task, we recorded their eye-movements and image selections.

3.2.4. Data processing & exclusion

The Tobii T60 eye-tracker had a sample rate of 60 frames per second. For each sample, gaze was categorized as falling in one of the four quadrants of the screen (each defined as 50% of the display height and width). Gaze coordinate estimates that did not fall on the screen were considered track loss. We omitted trials from our analysis that had excessive track loss (≥ 50% of samples) within the cluster analysis window (0–1500 ms after target word onset; see *§3.3.2.1. Cluster analysis*). We omitted any remaining samples of track loss prior to binning the data.

The data was processed into time bins of 50 ms. In the Q1 analysis, bins were defined relative to target word onset; in the Q2 analysis, bins were defined relative to the offset of the word after which the target becomes predictable in the constraining sentences (the *predictor word*; see *§3.4.2.1. Incremental cloze task*). An image's quadrant received a value of 1 (indicating a look) if ≥ 50% of recorded looks within that bin fell on the quadrant; otherwise, it received a 0.

We additionally recorded participants' image selections. Touch screen inputs were categorized as falling in one of the four quadrants of the screen (following the same method as the gaze location categorization); if the input fell in an image's quadrant, then the participant was considered to have selected that image.

In the primary sample, we omitted trials from our analysis in which participants did not correctly select the task image. In this sample, we also excluded participants with high levels of task inaccuracy (incorrect selections on ≥ 25% of trials; 9 children, 1 adult). Additional participants were excluded if trial omissions (due to inaccurate selections and/or track loss) resulted in loss of ≥ 50% of their experimental trials (6 child participants). In the primary child sample, the grand mean image selection accuracy was 93%; in the exploratory child sample, the grand mean image selection accuracy was 88%. In the primary adult sample, the grand mean image selection accuracy was 97%; in the exploratory adult sample, the grand mean image selection accuracy was 95%.

3.3. Q1: Use of top-down information during word recognition

The first question we sought to answer in our study was whether four and five-year-old children, like adults, can use top-down contextual cues to guide spoken word recognition,

allowing them to rule out semantically incongruent candidates. To answer this question, we tested whether the phonemic cohort effect (the difference in competitor image looks between the cohort and control conditions shortly after target word onset) differed in the constraining and neutral sentence conditions and whether this effect differed in our child and adult populations.

In the neutral sentences, when both the target and the competitor image name are plausible sentence continuations, we expect to see phonemic cohort effects, or a greater likelihood of looks to competitor images in the cohort condition (when their names match the onsets of the target words) than in the control condition (when they are unrelated to the target words). We expect to see this pattern for both populations; as mentioned in the Introduction, phonemic cohort effects have been previously observed in neutral sentence contexts for both adults (e.g., Allopenna et al., 1998; Dahan & Gaskell, 2007; Dahan et al., 2001; Farris-Trimble & McMurray, 2013; Magnusson et al., 1999; inter alia) and young children (e.g., Desroches et al., 2006; Kandel & Snedeker, 2024; Sekerina & Brooks, 2007; Rigler et al., 2015; Weighall et al., 2017; inter alia).

In the constraining sentences, the two populations could diverge. If participants use top-down cues to guide lexical processing, we would expect to see a reduction or absence of the phonemic cohort effect in the constraining sentences, reflecting the fact that participants are able to rule out the cohort competitor as a potential match to the input, even though it initially matches the bottom-up input. The size of this reduction may vary between young children and adults if young children are less adept at identifying and rapidly integrating relevant top-down cues. Alternatively, if sentence context does not guide lexical processing in young children, we would expect to see phonemic cohort effects in the constraining condition in this population, similar to those observed in the neutral condition.

Crucially, since the displays did not depict the target word (cf., Dahan & Tanenhaus, 2004), any reduction of the cohort effect we observe in the constraining sentence condition cannot result from anticipatory target looks drawing looks from the cohort competitor. By comparing looks to the same set of competitor images across conditions, we can be confident that any differences observed are not due to low-level properties of the images themselves or the words they represent.

3.3.1. Analyses

All analyses in the present study were conducted in R v4.1.0 (R Core Team, 2021). Models were fit using the {lme4} package v1.1-27.1 (Bates et al., 2015).

In order to address our question of interest, we tested whether the effect of competitor relatedness (i.e., the phonemic cohort effect) differed by sentence condition (neutral, constraining) and population (child, adult). We preregistered two types of analyses: time window analyses and cluster analyses. We additionally analyzed the influence of sentence condition on the phonemic cohort effect in the adult and child datasets separately; these individual dataset analyses produced a similar pattern of results to that reflected in the combined data analysis and are thus described in the Supplementary Materials.

3.3.1.1. Time window analyses

In the time window analyses, we assessed how the phonemic cohort effect varied by sentence condition and population within a defined time window. Given that the timing of phonemic cohort effects can vary between adults and children (Sekerina & Brooks, 2007), we used a collapsed localizer technique (e.g., Luck & Gaspelin, 2017) to identify our analysis time

window.[4] We collapsed the data across sentence conditions and populations and looked for an observed cluster for the effect of competitor relatedness (this clustering followed the same procedure described in *§3.3.1.1.2. Cluster analyses*; the analysis models computed at each step had a fixed effect of relatedness condition and random intercepts for participant and item). This observed cluster provides a sense of when a phonemic cohort effect is likely to arise in the data, independent of sentence condition and population (though see *§3.3.1.2. Cluster analysis* for discussion of cluster timing interpretation). We used the time bins in the observed cluster to define our analysis time window, thereby allowing us to assess how the phonemic cohort effect differs across our factors of interest despite potential differences in effect timing. The resulting time window was 550–949 ms after target onset.

To avoid concerns about time-related dependencies in our dependent variable (if a participant is looking in a location at one time bin, they are likely to be looking at the same location 50 ms later in the following time bin), we collapsed the data for each trial across the time window and binarized the measure of competitor image looks.[5] A trial received a 1 (indicating a look) if ≥ 50% of time bins within the window contained a look to the competitor image (analogous to the threshold in our binning procedure; see *2.4 Data processing & exclusion*); otherwise, the trial received a 0 (indicating no look). We conducted a 2 × 2 × 2

[4] Note that this time window selection procedure differs from our preregistration, which stated that we would use a time window defined by the earliest onset of the phonemic cohort effect in the neutral condition in the individual dataset cluster analyses and the offset of the phonemic cohort effect in the neutral condition for the analysis of the adult data. We decided to use a collapsed localizer technique instead in order to address potential concerns about false positives. The pattern of results is the same when using the preregistered time window (500–749 ms after target onset) and the localized time window (550–949 ms after target onset).

[5] Note that this collapsing procedure was not specified in our preregistered time window analyses (in either the Q1 or Q2 analyses). Analyzing the data while maintaining the time bin structure within the analysis window produces the same pattern of results (for this analysis, we added a random intercept of time bin to the preregistered model structure to account for time-related dependencies between bins).

analysis of competitor relatedness (control, cohort), sentence condition (neutral, constraining), and population (adult, child), assessing the likelihood of competitor image looks (0, 1) within the analysis time window. The analysis model had fixed effects of competitor relatedness condition, sentence condition, and population, with interactions between all three variables. For all analyses in the present study, fixed effect factors involved in interactions were entered into the models using effects coding (e.g., Hardy, 1993). The model had random intercepts for participant and item as well as random slopes for competitor relatedness condition, sentence condition, and their interaction by participant and item. In all analyses of the present study, item was defined by the target word. In this analysis, we were interested in the reliability of the two-way interaction between competitor relatedness and sentence condition as well as the three-way interaction with population, which provides an indication of whether the modulation of the phonemic cohort effect by sentence condition differed between the children and adults in our sample.

We additionally analyzed the data for the constraining and neutral sentences separately within the same time analysis window. For each sentence condition, we computed a model with fixed effects of competitor relatedness condition, population, and their interaction, random intercepts for participant and item as well as random slopes for cohort relatedness condition by participant and item. In these analyses, we were interested in the two-way interaction between competitor relatedness condition and population, which would indicate a difference in the size of the phonemic cohort effect between populations within a sentence condition.

3.3.1.2. Cluster analyses

We supported our time window analyses with cluster permutation analyses (e.g., Hahn et al., 2015; Slim et al., 2024; Yacovone et al., 2021). In the cluster analyses, we investigated

competitor image looks 0–1499 ms after target onset. These analyses had a step size of one time bin (50 ms). We first looked for sequential time bins in our data where the effect(s) of interest were reliable (the "observed cluster"). At each step, we computed a logistic mixed effects model to assess the effect(s) of interest. An effect was considered reliable at a step if the absolute value of its *z*-value was greater than 2 (Gelman & Hill, 2007).[6] A minimum of two adjacent steps with reliable effects was required to comprise a cluster. After identifying an observed cluster, we assessed its reliability using permutation. We performed 1000 simulations reshuffling the competitor relatedness condition and sentence condition trial labels for each participant and repeated the cluster analysis on the permuted data. For each simulation, for any identified clusters, we summed the absolute values of the *z*-values of the steps within the cluster to obtain a sum statistic (the *z*-sum), and we recorded the largest resulting sum statistic for each permutation. We compared the resulting distribution of sum statistics to the sum statistic of our observed cluster to determine its reliability; *p*-values for observed clusters were determined based on where its sum statistic fell in this distribution (e.g., a sum statistic ≤ 95% of the sum statistics in the distribution would have a *p*-value of < 0.05).

We looked for clusters of the three-way interaction between competitor relatedness condition, sentence condition, and population; the models computed at each step contained fixed effects of these three variables and their interaction as well as random intercepts for participant and item. We also looked for clusters of the two-way interaction between competitor relatedness and population within the neutral sentences and constraining sentences separately; the models

[6] Following Yacovone et al. (2021), if a model failed to converge at a step (excluding singular fit warnings), instead of using the computed model estimates for that step, we used the model estimates from the prior step (if this occurred at the first step, the model estimates were set to zero). This procedure prevents models that do not converge from prematurely ending or otherwise breaking up a cluster.

computed at each step contained fixed effects of these two variables and their interaction as well as random intercepts for participant and item.

Note that while cluster permutation analyses can provide information about the presence of effects and approximately when they arise, these analyses cannot be used to make precise inferences about the onset, offset, and durations of effects (for discussion, see Fields & Kuperberg, 2020; Groppe et al., 2011; Sassenhagen & Draschkow, 2019; Slim et al., 2024).

3.3.2. Results

Figure 3.2 plots grand mean looks to the competitor image over time in the child and adult data (see Supplementary Materials for a plot of task image looks). The pattern of looks is visually similar in both populations. For both children and adults, looks to the competitor image began roughly at chance (25%) in both the cohort and control conditions in both the neutral and constraining sentence conditions. In the neutral sentences, looks to the competitor image increased briefly after target onset in the cohort condition, suggesting that participants initially considered the cohort competitor to be a potential match to the input. In the constraining sentences, looks to the competitor image remained roughly at chance in both the cohort and control conditions, implying that participants did not consider either image as a potential match to the input as they heard the target word.

This pattern was supported by our analyses (Figure 3.3). In the time window analysis (550–949 ms after target onset), there was a two-way interaction between competitor relatedness condition and sentence condition, indicating that there was a reliable difference in the size of the phonemic cohort effect between sentence conditions (β = -0.103, SE = 0.048, z = -2.162, p = 0.031). The phonemic cohort effect was reliable in the analysis of the neutral sentences (β =

0.262, SE = 0.078, z = 3.378, p < 0.001) but not in the constraining sentences (β = 0.031, SE = 0.072, z = 0.426, p = 0.670). However, the three-way interaction between competitor relatedness condition, sentence condition, and population was not reliable (β = 0.013, SE = 0.045, z = 0.287, p = 0.774), and there were no reliable two-way interactions between competitor relatedness condition and population in either the neutral sentences (β = -0.041, SE = 0.063, z = -0.658, p = 0.511) or constraining sentences (β = -0.012, SE = 0.063, z = -0.193, p = 0.847), suggesting that the modulation of the phonemic cohort effect by sentence condition did not reliably differ between children and adults. The cluster analyses identified a trending cluster with a two-way interaction between competitor relatedness and sentence condition (550–749 ms after target onset; z-sum = 9.86, p = 0.069), however there were no time bin clusters with a three-way interaction suggesting a modulation of this cohort effect by population, and there were no clusters in either sentence condition with an interaction between competitor relatedness and population.

Figure 3.2: *Grand mean looks to the competitor images in the cohort and control conditions by sentence condition and population.* Ribbons indicate standard error. Lines indicate reliable clusters of the effect of competitor relatedness in the individual dataset analyses (see Supplementary Materials for details).

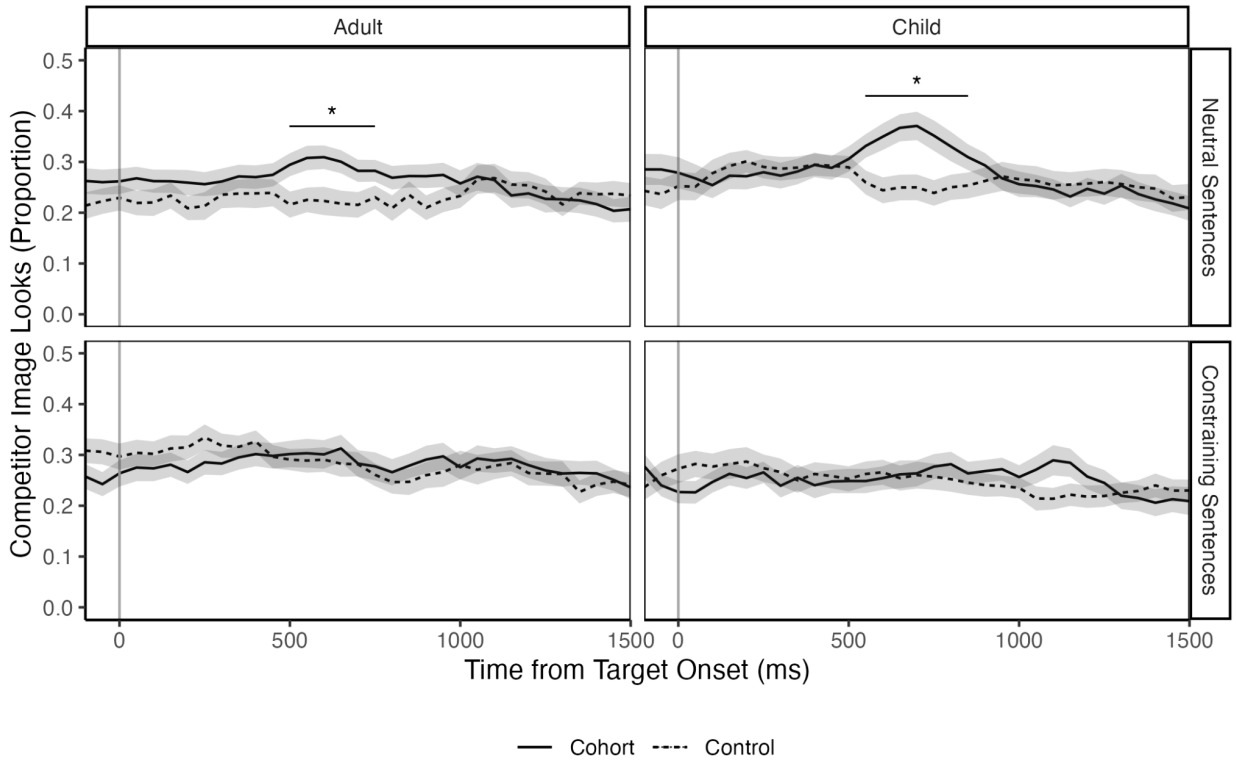
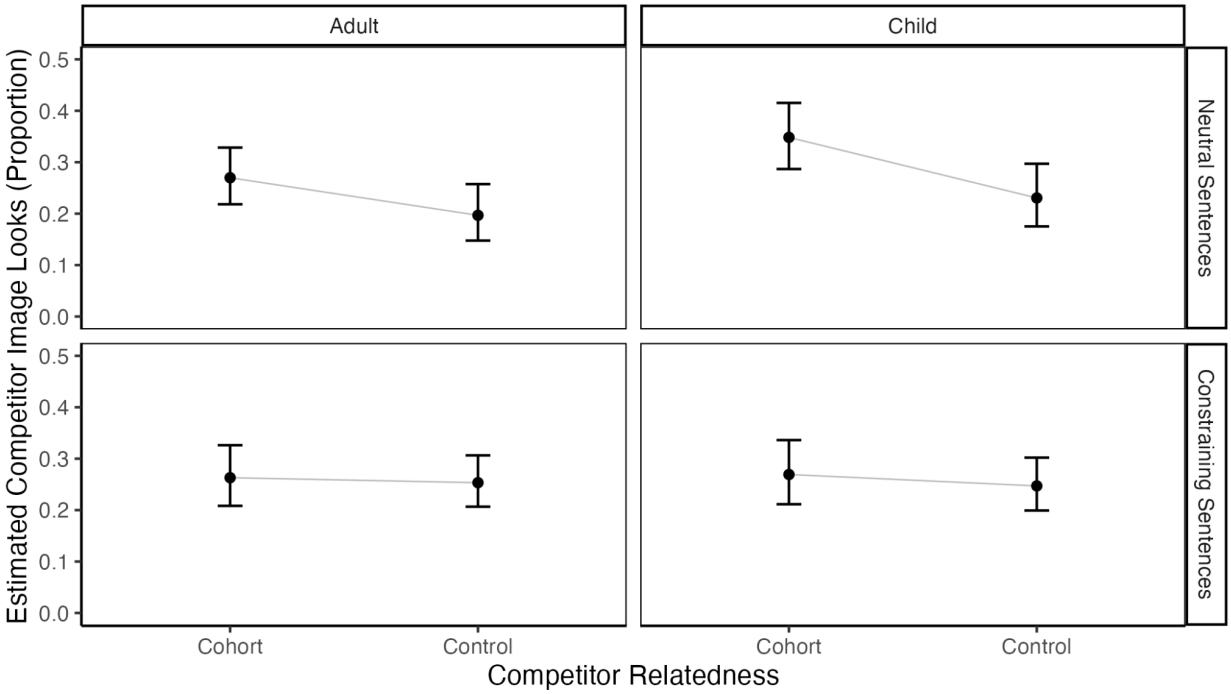


Figure 3.3: *Competitor relatedness × sentence condition × population effect plot within the time window 550–949 ms after target onset.* Error bars indicate 95% confidence intervals.



3.3.3. Child image selections

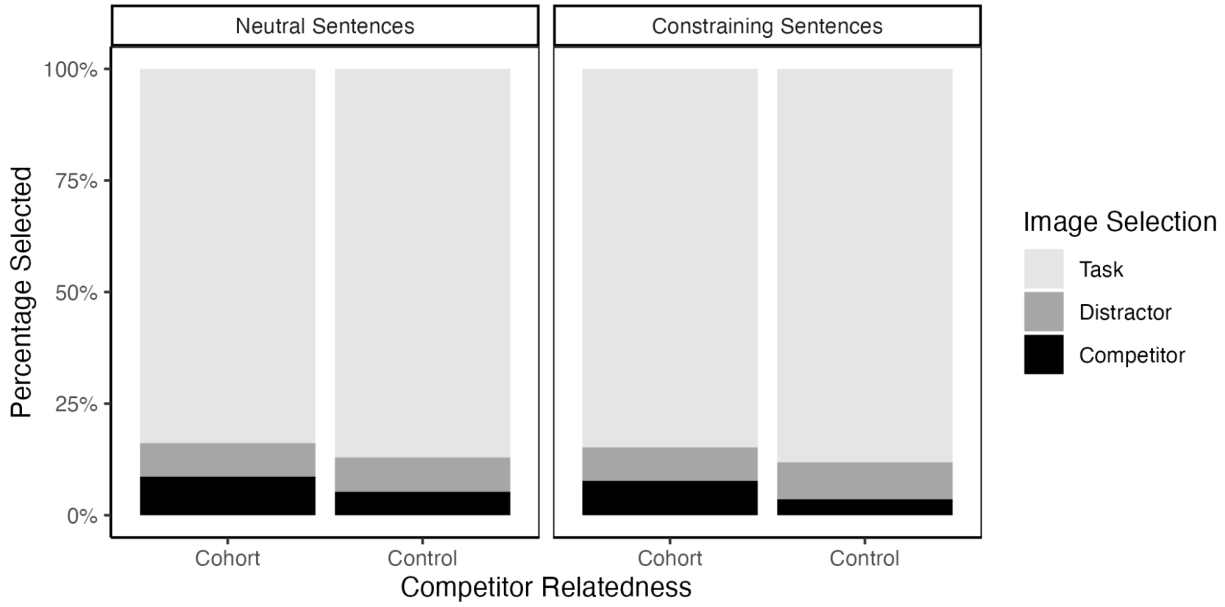
Interestingly, we observed additional evidence of phonemic cohort competition in the children's selection data. Child participants made many more incorrect image selections than adults, with 9 children omitted from our primary sample due to high levels of inaccuracy (incorrect selections on ≥ 25% of trials). Figure 3.4 shows the distribution of incorrect selections in the experimental trials made by all children who completed the task (N = 71; M_{age} = 4.9 years, SD = 0.5, range = 4;0;12–5;9;29, 37 F, 34 M; though note that we also observe the same pattern of results in our primary sample). Errors were made on 12% of experimental trials.

In post-hoc analyses, we tested the likelihood of mistaken competitor image selections (0,1) and mistaken distractor image selections (0,1) using logistic mixed effects models with

fixed effects of competitor relatedness, sentence condition, and their interaction along with random intercepts for subject and item.

Although we observed no differences in the likelihood of mistaken distractor image selections across competitor relatedness conditions (β = -0.045, SE = 0.088, z = -0.516, p = 0.606), participants were significantly more likely to mistakenly select the competitor image in the cohort condition than the control condition (β = 0.353, SE = 0.102, z = 3.464, $p < 0.001$), suggesting that the children in our experiment were at times unable to inhibit the competition from phonemic cohort competitors. Surprisingly, this pattern appeared in both sentence conditions (the competitor relatedness × sentence condition interaction was not reliable; β = 0.042, SE = 0.101, z = 0.410, p = 0.682). In contrast, adults made fewer errors overall, and errors did not vary reliably across conditions.

Figure 3.4: *Distribution of image selections by image type across competitor relatedness and sentence conditions.*



3.3.4. Q1 summary

In our Q1 analyses, we observed evidence of phonemic cohort competition in neutral sentence contexts but not in constraining sentence contexts, suggesting that participants used top-down contextual cues when available to avoid competition from incongruent lexical candidates. This modulation was significant and did not reliably differ between the children and adults in our sample. These results thus demonstrate that by four to five years of age, children are able to use contextual information to constrain word recognition to approximately the same degree as adults, providing evidence that top-down language comprehension pathways are robust and active prior to formal schooling and literacy. While these findings contrast with prior child studies showing reduced or no integration of top-down information in a similar or older age range (e.g., Joseph et al., 2008; Khanna & Boland, 2010; Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008; Tiffin-Richards & Schroeder, 2020; Trueswell et al., 1999; Yacovone et al., 2021), they align

with recent child EEG work showing evidence of top-down lexical prediction in naturalistic contexts (Levari & Snedeker, 2024). We address potential reasons for this contrast with prior work in the General discussion.

Interestingly, the children in our sample at times selected the competitor image instead of the task word image and did so reliably more in the cohort condition than the control condition, suggesting that cohort competitor representations were activated by the input and that children were sometimes unable to inhibit this activation. This finding aligns with previous results illustrating that young children have more difficulty inhibiting lexical representations once they have been activated, resulting in perseveration of competition effects and/or incorrect selections of competitors (e.g., Huang & Snedeker, 2011; Sekerina & Brooks, 2007). Surprisingly, we observed this pattern in both the neutral and constraining sentence conditions, even though participants only showed a phonemic cohort effect in the eye-tracking data in the neutral sentences (participants did not look to cohort competitors when hearing the target word in the constraining sentences). These results suggest that although children were able to avoid competition from the cohort competitor during target word recognition in the constraining sentences, cohort competitor representations were (at least at times) still activated by the sentence. Activation of the cohort competitor could arise in the constraining sentences if participants use the available contextual cues to predict the target word and pre-activate its phonological form, which in turn would activate the phonological representation of the cohort competitor. In the next section, we investigate whether the participants in our experiment made form-based predictions in the constraining sentences.

3.4. Q2: Phonological prediction in constraining sentences

The second question we addressed in our study was whether participants could use the contextual information in the constraining sentences to predict upcoming target words to the level of phonological form. As mentioned in the Introduction, prior work with adults has found evidence of predictive phonological competitor effects, or early phonemic cohort effects arising prior to the onset of the target word (see Ito, 2024 for review). These early effects suggest that the phonological form of the target word has been activated before it is encountered in the input, and this activation drives fixations on the competitor word image given the overlap in phonology.

We investigated whether the participants in our experiment similarly display predictive phonological competitor effects in the constraining sentences, when there are cues available for prediction. However, our analysis approached this effect in a different way from prior work. In virtually all prior studies showing this effect, competitor image looks were analyzed relative to target word onset, with the visual display appearing a fixed amount of time prior to target onset (typically 1000 or 2000 ms) (Ito, 2024, Table 3.1; cf., Kukona, 2020). This design allows for comparability across sentences with different target onset times, since the onset of the predictive phonological competitor effect is restricted by the onset of the visual display (it is not possible to launch competitor image looks before the competitor image appears on the screen), which in turn is time-locked to target onset. In fact, Li et al. (2022) tested two different display onset times (1000 ms and 2000 ms prior to target onset) and found that the time-course of predictive phonological competitor effects is to be linked to the onset of the visual display, with effects arising approximately 600 ms after display onset.

Our experiment did not follow this same design; the images were on screen for 4000 ms prior to sentence onset (2000 ms of preview time, 2000 ms of central fixation). Under the hypothesis that participants look to phonemic cohort competitors after activating the phonological form of the target word, we should expect competitor image looks to increase in the cohort condition once the target word becomes predictable in our stimulus sentences and participants have had time to activate its phonological form. This point occurs at different times relative to sentence and target word onset across our stimulus sentences, given that they varied in structure (see Supplementary Materials for stimulus sentences). Consequently, we preregistered an exploratory analysis looking for evidence of early phonemic cohort effects arising shortly after target words become predictable in the sentences (as identified via an incremental cloze task) but before they are heard in the input.[7] This analysis provides a direct test of predictive phonological competition that doesn't rely on controlling the visual world display onset time.

To address this question, we analyzed the exploratory sample (see *§3.2.1. Participants*), which included more participants due to eliminating the accuracy-based exclusion criteria from our pre-registration. This decision was motivated by the observation that the child participants in our experiment at times selected cohort competitors in place of the task word, potentially reflecting a reduced ability to inhibit cohort competition (see *§3.3.3. Child image selections*). We observed this pattern in the constraining sentences as well as the neutral sentences, even though there was no evidence of cohort competition in the constraining sentences as participants heard the target word. The participants and trials for which cohort competition was so great in the constraining sentences that it led to incorrect selections may be precisely those for which

[7] In our initial preregistrations of the adult and child dataset analyses, we included exploratory analyses investigating competitor image looks in a two second window prior to target word onset. These analyses are described in the Supplementary Materials, as the preregistered exploratory analysis described here provides a more direct test of our phenomenon of interest.

phonological prediction was most likely to have occurred. Consequently, we thought it would be informative to include the eye-movements from these participants and trials in the Q2 analysis. In order to avoid inflating the likelihood of false positives in our analysis, we took a conservative approach and included all trials with incorrect selections and did not omit any participants for reasons of task inaccuracy. In the present section, we report the results from the analysis of the exploratory sample and identify in footnotes the instances when the pattern of the results differed for the primary sample. Note that the overall pattern of Q1 results was the same in the exploratory sample.

3.4.1. Analysis

3.4.1.1. Incremental cloze task

We conducted an incremental cloze task to determine the point in each constraining sentence when the target word becomes predictable. We trimmed our audio stimuli for the constraining sentences to create sentence fragments, ranging in length from the first word of the sentence to the length of the sentence truncated two words prior to target onset (Table 3.3) (cloze information one word prior to target onset was derived from our norming task; *§3.2.2.1. Sentence-level controls*). In the cloze task, participants listened to these sentence fragments and were instructed to type them and complete them as full sentences.

Participants in this task were adult native, monolingual speakers of American English (recruited via Prolific). Participants were presented with one fragment per constraining sentence. We recorded at least $n = 20$ responses per fragment (since the sentences varied in length, there were a different number of fragments per sentence, and some fragments received more responses

than others; M = 42 responses, SD = 18, range = 20–100). Items were balanced across the different audio versions of the constraining sentences (see *§3.2.2.2. Audio splicing*). For each response, we recorded whether or not it contained the target word. For each sentence, we identified the first word after which ≥ 50% of responses included the target (the *predictor word*).

On average, the predictor word was positioned four words prior to the target word (SD = 3, range = 1–11). Any items in which the predictor word was only one word prior to the target were omitted from analysis (n = 6). We re-binned the eye-tracking data relative to the predictor word offset. We analyzed this data using both a time window analysis and a cluster permutation analysis.

Table 3.3: *Example stimuli and responses in the incremental cloze task for the constraining sentence "The baby drank the* ***milk****…".* In this item, the predictor word was *drank*.

Sentence fragment	Example response	Responses containing the target word
The	The pig flew	0%
The baby	The baby is asleep	3%
The baby drank	The baby drank the milk from the bottle	63%

3.4.1.1. Time window analysis

In the time window analysis, we analyzed looks to the competitor images in a pre-registered time window extending from 0–499 ms after predictor word offset. This window duration was selected such that it would be long enough to include a predictive phonological competitor effect of the duration observed by Ito et al. (2018) (150 ms) or Li et al. (2022) (300 ms), while also allowing 200 ms for saccade execution to the competitor image (e.g., Allopenna

et al., 1998; Cooper, 1974). Note, however, that this window assumes analogous effect onsets when measured from the point of target word predictability in our paradigm and from visual display onset time in Ito et al.'s (2018) and Li et al.'s (2022) paradigms. There may be reasons to expect differences in effect timing between paradigms and for our child participants, for whom prediction may take longer than adults. We address this uncertainty in effect timing by additionally conducting a cluster analysis (*§3.4.1.2*).

Looks within the preregistered time window were analyzed using logistic mixed effects models. As in the Q1 analyses, we collapsed the data for each trial across the time window and binarized the measure of competitor image looks (see *§3.3.1.1*). The adult and child data were analyzed separately. If the target word onset for a sentence was less than 500 ms from the predictor word offset, then we only included the time bins prior to the one containing the target word. If this resulted in a window for that sentence that was less than 200 ms long, the item was dropped from the analysis (n = 3). This left 21 items in the time window analysis.

The analysis models contained a fixed effect of competitor relatedness condition, random intercepts for participant and item as well as random slopes for competitor relatedness condition by participant and item. We additionally ran a post-hoc time window analysis to account for the fact that the exploratory sample included trials with incorrect image selections. This model had the same structure as that described above with the addition of a fixed effect of selection accuracy (correct, incorrect) and its interaction with competitor relatedness condition.

3.4.1.2. Cluster analysis

The cluster analysis procedure followed the same format as in the Q1 analyses (see *§3.3.1.2. Cluster analyses*). The models computed at each step of the analysis had a fixed effect

of competitor relatedness condition and random intercepts for participant and item. The cluster analysis window extended from 0 ms after predictor word offset to no further than target word onset. This analysis included the $n = 24$ items whose predictor word did not directly precede the target. As our sentences had different durations between the offset of the predictor word and the onset of the target word, the number of data points available for analysis decreased in later time bins. To avoid potential issues of sparsity, we only analyzed time bins that had at least 100 data points. The resulting cluster analysis time windows were 0–2199 ms after predictor word offset in both the adult and child analyses.

3.4.2. Results

Figure 3.5 plots the grand mean looks to the competitor image in the constraining sentences over time, relative to predictor word offset, in both the adult and child samples (see Supplementary Materials for plots showing looks in the neutral sentences). In the adult data, competitor image looks remain similar in both competitor relatedness conditions. In the child data, competitor image looks begin to increase in the cohort condition starting between predictor word offset and average target word onset.

Our analysis confirms the veracity of these observations. In the adult data, the effect of competitor relatedness was not reliable in the preregistered time window analysis ($\beta = -0.151$, SE = 0.219, $z = -0.688$, $p = 0.492$) or the post-hoc analysis accounting for selection accuracy ($\beta = 0.044$, SE = 0.224, $z = 0.199$, $p = 0.842$), and there were no identified phonemic cohort effect clusters. In the child data, although there was no reliable phonemic cohort effect in the preregistered time window analysis ($\beta = -0.383$, SE = 0.285, $z = -1.343$, $p = 0.179$), there was a significant overall effect of competitor relatedness in the post-hoc analysis accounting for

selection accuracy (β = 0.463, SE = 0.178, z = 2.608, p = 0.009). In fact, there was a significant interaction between competitor relatedness and selection accuracy such that the phonemic cohort effect was larger in trials in which an incorrect selection was made (β = -0.382, SE = 0.149, z = -2.567, p = 0.010). In the cluster analysis, there were reliably more looks to competitors in the cohort condition than the control condition from 250–899 ms after predictor word offset (z-sum = 38.117, p = 0.016).[8] Given that this predictive phonological competitor cluster arose much later than our preregistered time window (0–499 ms after predictor word offset), we ran additional, post-hoc time window analyses using the cluster time window. In this window, we observed reliable early phonemic cohort effects both with the preregistered model structure (β = -0.757, SE = 0.217, z = -3.485, p < 0.001) and when accounting for selection accuracy (β = 0.685, SE = 0.168, z = 4.091, p < 0.0001). As in the preregistered time window, the cohort effect was larger in trials in which an incorrect selection was made (β = -0.403, SE = 0.162; -2.491; p = 0.013; Figure 3.6).

[8] There was additionally a later, trending cluster from 1200–1450 ms after predictor word offset (z-sum = 14.621, p = 0.076). No clusters were identified in the analysis of the primary sample.

Figure 3.5: *Grand mean early looks to the competitor images in the constraining sentences after predictor word offset, broken down by competitor relatedness and population.* Ribbons indicate standard error. Lines indicate reliable clusters of the effect of competitor relatedness; the cluster analysis did not include samples after target word onset.

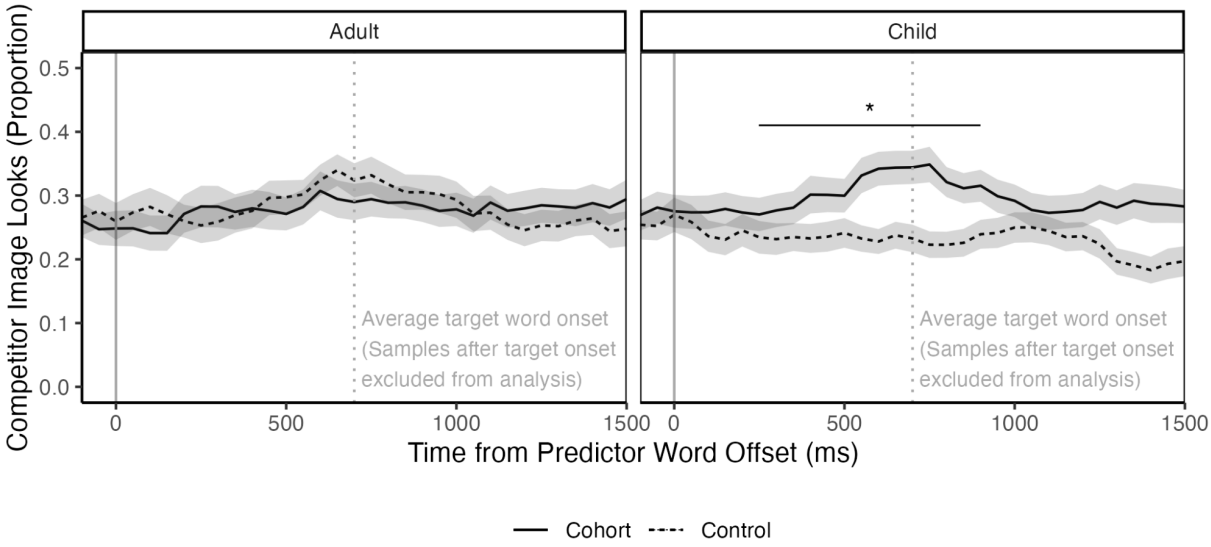
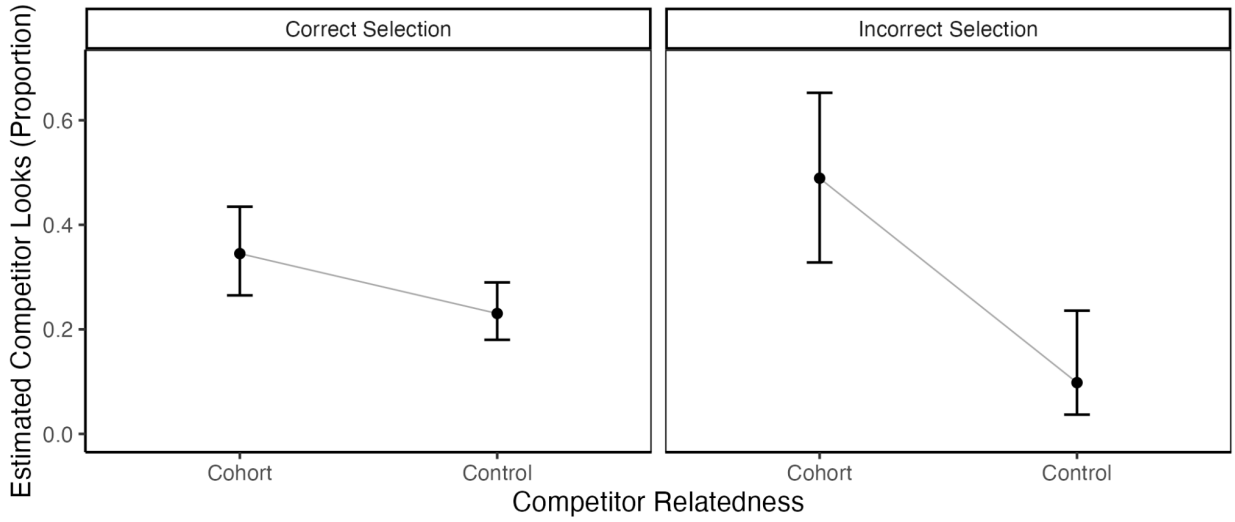


Figure 3.6: *Competitor relatedness × selection accuracy effect plot in the child data within the post-hoc time window 250–899 ms after predictor word offset (excluding samples after target word onset).* Error bars indicate 95% confidence intervals.



3.4.3. Q2 summary

In our Q2 analysis, we observed predictive phonological competition in the eye-movements of four and five-year-olds: In the constraining sentences, child participants looked to images of phonemic cohort competitors shortly after the target words became predictable in the sentences but before they were articulated. Given this timing, these early looks cannot be explained by the same mechanism as cohort competition effects occurring shortly after target word onset (as in the neutral sentences in Q1), which reflect competition during decoding of the input. Instead, these results suggest that children were able to use the available contextual cues in the constraining sentences to predict the target word and pre-activate its phonological form, which then drove looks to cohort competitor images due to their phonological overlap. This pre-activation may explain why the children in our experiment made so many mistaken cohort competitor image selections in the constraining sentences (see *§3.3 Child image selections*), even

though they showed no evidence of cohort competition when hearing the target word (see Q1 above): The cohort competitor was activated in these sentences prior to target word onset due to predictive processing. In fact, the predictive phonological competitor effect was stronger for trials in which children made an incorrect selection.

In general, however, it appears that any activation of the cohort competitors resulting from predictive processing is temporary; by target word onset, children were not looking at the competitor image in the constraining sentences (Figure 2). One possibility is that the early activation of cohort competitors resembles the process of word planning during language production, in which phonologically-related words may become active before the correct target is produced (e.g., Dell, 1986; Dell & Reich, 1981; Harley, 1984). In fact, it has been hypothesized that production mechanisms may at times play a role in linguistic prediction (Dell & Chang, 2014; Federmeier, 2007; Lelonkiewicz et al., 2021; Martin et al., 2018; Pickering & Gambi, 2018; Pickering & Garrod, 2007, 2013a).

Surprisingly, given prior work with adults (see Ito, 2024), we did not observe a comparable predictive phonological competitor effect in our adult sample. However, since the child participants in our study showed evidence of phonological prediction, we do not consider this null finding to be an indication that adults are unable to perform form-based prediction or that the contextual cues in our sentences were not sufficiently constraining for such predictions (see Ito, 2024 for evidence that the size of the predictive phonological competitor effect is modulated by target word predictability). The contrast between the child and adult results may instead reflect that adults in our sample were more adept at inhibiting cohort competitor activation. Our task design encouraged participants to continuously scan the image display as they heard the stimulus sentence in order to identify the task image. It could be the case that

while some children struggled to inhibit the lexical representation of the cohort competitor once it had been activated, leading their gaze to linger on cohort competitor images, adults were able to rapidly inhibit any cohort competitor activation caused by prediction of the target word phonology in order to continue searching the display.

The contrast between our null finding with adults and prior work may also be explained by the fact that the images were onscreen for participants to search throughout the full stimulus sentence. In prior tasks (e.g., Ito et al., 2018; Li et al., 2022), the visual display appeared mid-sentence, at a point when the target word may have already been predicted. The appearance of the display likely prompted participants to identify and activate the names of the images on screen. It is possible that this activation of the cohort competitor phonology, coupled with the overlapping, additional activation from the predicted target word phonology, may have been more difficult for adults to inhibit than activation from the predicted target word phonology alone (as would be experienced in our task). Indeed, Ito and Husband (2017) similarly did not observe a predictive phonological competitor effect when the visual display appeared 3000 ms prior to sentence onset rather than shortly before target word onset as in Ito et al. (2018), inter alia.[9] Future research should continue to investigate what factors influence how and when predictive phonological competitor effects arise in eye-tracking studies. The fact that we observed an effect with children but not with adults should additionally prompt researchers to carefully consider how to interpret the absence of predictive phonological competition effects in

[9] While Kukona (2020) did observe a predictive phonological competitor effect with the visual display appearing prior to sentence onset, the target image was included in the display with the cohort competitor, meaning that the target image may have been referentially predicted (e.g., Altmann & Kamide, 1999), prompting activation to spread to the phonological competitor (see Ito, 2024 for discussion).

a given task and whether null findings should be interpreted as an indication that participants did not predict phonological form.

3.5. General discussion

In the present study, we investigated whether the top-down pathways necessary for interactive processing during language comprehension are in place by early childhood. We observed evidence of interactive processing in four and five-year-old children in two measures:

First, we observed that young children, like adults, are able to use top-down contextual information to guide bottom-up processing of spoken language input during word recognition (Q1). The participants in our study were able to use the cues from constraining sentence contexts in order to avoid competition from lexical candidates that partially matched the bottom-up input but were semantically incongruent in the context, resulting in a modulation of the phonemic cohort effect based on target word predictability (constrained vs. neutral). This ability to constrain bottom-up processing in real-time based on top-down information arose in both our adult and child samples, with no reliable differences between populations.

Second, we found evidence for form-based prediction in four and five-year-old children (Q2). In constraining sentences containing predictable target words, the children in our study looked to phonemic cohort competitors of these words shortly after the targets became predictable in the sentences, resulting in a predictive phonological competitor effect (Ito, 2024). This effect suggests that children predictively pre-activated the target word's phonology, which then drove looks to the cohort competitor image due to their shared phonology. We did not observe this same effect in our adult sample, likely reflecting the fact that young children are more susceptible to competition than adults — indeed, the children in our study were so

susceptible to competition that they at times incorrectly selected cohort competitor images in the experiment task (see Huang & Snedeker, 2011 for a similar effect).

These results suggest that the top-down pathways involved in language comprehension are robust and active early, prior to formal schooling and literacy. The early availability of these pathways could indicate that they are a fundamental consequence of the mind's architecture. In the remainder of the General discussion, we (i) consider (and reject) the possibility that these effects reflect lateral priming rather than top-down activation, (ii) place our findings in the context of prior work investigating the integration of top-down information in young children's language comprehension, and (iii) discuss how the early availability of top-down information during language comprehension may influence language development.

3.5.1. The source of target word form pre-activation: Top-down vs. spreading activation

The results of the present study offer insight into when and how word representations are activated during young children's real-time language processing. Our findings suggest that in the absence of contextual cues (as in the neutral sentences), word forms are activated primarily by the bottom-up input. In contrast, when words appear in sufficiently supportive contexts (as in the constraining sentences), their forms are activated in advance by the sentence context. This predictive pre-activation has the potential to constrain later processing, preventing the consideration of other lexical candidates when the word form is later encountered in the bottom-up input (as we observed in the Q1 analyses).

There are two possible, non-mutually exclusive routes by which children may pre-activate target word forms in constraining sentences. One possibility is that this activation arrives in a top-down fashion, cascading from higher-level representations. For example, when hearing

an utterance such as *The baby drank…*, participants may use their world knowledge to predict that the object of the baby's drinking action will be milk. The activation of *milk*'s conceptual representation may activate the corresponding lexical representation, which in turn activates its phonological representation, resulting in pre-activation of the target word form. Alternatively, activation may spread laterally to the target word representation from the word representations that are activated by the sentence input prior to target onset. For example, when the input activates the lexical representation *baby* in long-term memory, activation may spread from this representation to the stored representations of related words, including that for *milk*. The activation of *milk*'s lexical representation may then cascade to its phonological form. There is evidence that both top-down and passive spreading activation can lead to pre-activation of word representations, depending on context (Lau et al., 2013).

To test whether the eye-tracking effects we observed with our child participants reflect top-down processing, we performed post-hoc analyses to disentangle the contributions of top-down activation and spreading activation (see Supplementary Materials for details and full results). A target word's top-down activation was operationalized as the word's cloze probability within the relevant context. This measure reflects whether top-down processing leads to the generation of the target word based on the given contextual information in a cloze production task. A target's spreading activation was represented by the semantic relationship between the target word and the relevant context, operationalized by LSA (Landauer et al., 1998; Wolfe & Goldman, 2003).

To assess the role of top-down and spreading activation in children's word recognition (the Q1 effect), we investigated how looks to the competitor images in the cohort and control conditions were affected by the target words' LSA and cloze within their sentence contexts. We

observed a reliable influence of target word cloze on competitor image looks above and beyond the influence of LSA. As target word cloze increased, the likelihood of looks to cohort competitors — but not control competitors — decreased, meaning that the phonemic cohort was smaller when target word cloze was higher (as reflected by a reliable interaction between competitor image relatedness and target word cloze). These results illustrate the same difference between the neutral and constraining sentence conditions that we observed in the Q1 analyses, indicating that young children were better able to avoid competition from cohort competitors during word recognition when the context pointed to an alternative, highly predictable word. In contrast, target word LSA did not have a reliable effect on competitor image looks. This contrast suggests that top-down activation played a larger role than spreading activation in children's use of context to guide word recognition.

To assess children's pre-activation of phonological forms in the constraining sentences (the Q2 effect), we identified the target words' LSA and cloze at the point in the sentences when the target word first becomes predictable (see *4.1.1 Incremental cloze task*; we omitted items for which this sentence point was only one word prior to the target). Neither measure had a reliable influence on competitor image looks in the time window where the predictive phonological competitor effect arose in the child data, meaning we are unable to conclude from this analysis that target word cloze had an effect above and beyond that of target LSA. Nevertheless, when we analyzed competitor image looks for the subset of items in which target words had high cloze but low LSA (< 0.25), we still observed a predictive phonological competitor effect in this window, suggesting that target word form pre-activation can arise even when the target word does not have a particularly high semantic association with the preceding context. Consequently, while we

cannot rule out that spreading activation contributed to the Q2 effect, the effect appears not to rely solely on spreading activation, suggesting a role for top-down activation.

Taken together, these results support the interpretation of our data patterns as reflecting top-down processing in the language comprehension of four and five-year-old children (in addition to any passive spreading activation). This top-down processing may be surprising, given prior work showing limited or no integration of top-down information during sentence processing in children of similar and even older ages (e.g., Joseph et al., 2008; Khanna & Boland, 2010; Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008; Tiffin-Richards & Schroeder, 2020; Trueswell et al., 1999; Yacovone et al., 2021). We address this contrast in the following section.

3.5.2. Contextualizing our findings with prior literature on child language comprehension

The present set of results suggest that the pathways necessary for top-down processing in language comprehension are already in place by the fourth and fifth year of life. The results of the Q1 analyses are consistent with eye-tracking work suggesting that children (6–9 years) are able to use sentence context to avoid activating incongruent homophone meanings during spoken word recognition (Hahn et al., 2015). The results of the Q2 analyses complement and extend findings that the ability to anticipate upcoming referents arises early (e.g., Borovsky et al., 2012; Borovsky et al., 2014; Lukyanenko & Fisher, 2016; Mani & Huettig, 2012; Nation et al., 2003; Özge et al., 2019; Sommerfeld et al., 2023; inter alia), illustrating that young children can additionally anticipate and pre-activate word representations corresponding to these upcoming referents, including the words' phonological forms. These results are also consistent with recent EEG work showing that school-aged children make predictions about upcoming words during

naturalistic comprehension (Levari & Snedeker, 2024) as well as cloze production work showing that five and six-year-old children can generate explicit word predictions (Waite et al., under review).

However, the study findings contrast with another body of work showing limited or no real-time integration of top-down information in young children. As mentioned in the Introduction, young children appear less adept than adults at using top-down information to resolve ambiguity during syntactic parsing (e.g., Kidd et al., 2011; Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008; Trueswell et al., 1999; Yacovone et al., 2021). In particular, when young children encounter an utterance that is structurally ambiguous, as in the sentence *Tickle the frog with the fan*, where the prepositional phrase *with the fan* could either modify the noun *frog* or the verb *tickle*, they tend to base their interpretations on bottom-up cues such as intonation and lexical biases (e.g., whether verbs tend to co-occur with instruments), ignoring potential top-down cues such as plausibility (Are fans likely to be used for tickling?) or referential context (Is the speaker trying to distinguish between a frog with a fan and a frog without a fan?). This body of work primarily focuses on children of the same age as the present study (4–5 years). Other studies, however, have found that even older children (approximately 6–9 years) are less proficient than adults at integrating top-down information into lexical and syntactic processing (e.g., Joseph et al., 2008; Khanna & Boland, 2010; Tiffin-Richards & Schroeder, 2020). We present a few potential, non-mutually exclusive explanations for this contrast.

First, the contrast may reflect differences in demands imposed by the experimental task. For example, all of the studies illustrating that school-aged children show reduced top-down integration in lexical processing (Khanna & Boland, 2010; Joseph et al., 2008; Tiffin-Richards &

Schroeder, 2020) employ tasks that involve reading. Most seven to nine-year-olds have only recently learned to read and thus may have to attend more to decoding the bottom up input, leaving them with limited resources for top-down processing. Indeed, while school-aged children failed to use context to constrain homophone recognition in Khanna and Boland's (2010) cross-modal reading paradigm, they succeeded in Hahn et al.'s (2015) visual world paradigm with spoken language.

In the remainder of the studies in which children showed limited or no top-down integration (e.g., Kidd et al., 2011; Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008; Trueswell et al., 1999; Yacovone et al., 2021), successful comprehension required resolving syntactic ambiguity (specifically, prepositional phrase attachment ambiguity). Syntactic ambiguity resolution may be more difficult for young children than recognizing and predicting unambiguous, known words during spoken language comprehension, as was the measure of top-down processing in the present study. For instance, correctly resolving the syntactic ambiguity in a sentence such as *Cut the cake with the candle* (Kidd et al., 2011) requires the listener either (i) to hold multiple possible interpretations in mind during processing (*with* modifying *cut* vs. *with* modifying *cake*), shifting between these interpretations to assess how well they match the incoming input until they receive sufficiently disambiguating information to select one over the other (e.g., the word *candle*, which is an unlikely instrument for cutting) or (ii) to select an initial parse (e.g., driven by bottom-up information such as the fact that *cut* often co-occurs with an instrument modifier) that must be overridden and revised if the listener encounters later information that is incongruent with this parse (e.g., the word *candle*). Age-related differences in working memory capacity (e.g., Chi, 1978; Dempster, 1981; Schneider & Bjorklund, 1998), executive functioning (e.g., Best & Miller, 2010; Mazuka et al., 2009), and inhibitory control

(e.g., Carver et al., 2001; Johnstone et al., 2005; Passler et al., 1985; Welsh et al., 1991; Wiebe et al., 2012) would make either of these processes difficult for young children. Thus, in cases of syntactic ambiguity, children may commit to an initial parse based on the most valid and accessible interpretation cues that they then fail to revise (see Trueswell et al., 1999) — if the most valid and accessible cues to young children are bottom-up cues rather than top-down ones (see below), then this would lead to limited evidence of top-down interactivity.

Next, the contrast may reflect differences in the validity of the available top-down cues. In many studies of prepositional phrase attachment ambiguity in children, the available top-down cue to the intended sentence parse is provided by the referential context (e.g., Snedeker & Trueswell, 2004; Trueswell et al., 1999). Adults can use information about the potential referents in a context to constrain their interpretation of sentences (e.g., Altmann & Steedman, 1988; Tanenhaus et al., 1995; Trueswell et al., 1999). For instance, if adults are instructed to *Tickle the frog with the fan*, if there is a single frog in the referential context, they interpret the prepositional phrase *with the fan* as modifying the verb *tickle* (i.e. *Tickle the frog by using the fan*); if there are two frogs in the context (one with a fan and one without a fan), adults are more likely to interpret the prepositional phrase as modifying the noun *frog* (i.e., *Tickle the frog that has a fan*) (Snedeker & Trueswell, 2004). This behavior suggests that adults have recognized a contingency between the presence of referential ambiguity (resulting from multiple possible referents for a noun) and the likelihood of hearing a *with*-phrase that disambiguates the referent. However, the presence of multiple possible referents for the noun *frog* only weakly predicts the likelihood of a syntactic structure involving noun modification (adults often produce under-specified descriptions of nouns; Brown-Schmidt et al., 2002). This reduced validity could make it more difficult for young children to acquire the contingency between a speaker's referential

model and the likelihood of referent disambiguation with a *wh*-phrase that is necessary to use referential context as a top-down cue, leading young children to rely instead on more reliable, bottom-up cues to syntactic structure such as verb biases or prosody when interpreting *wh*-phrases. In contrast, in the present study, the target words in the constraining sentences are highly predictable from contextual cues (average target word cloze was 75%), making the available contextual information a robust and reliable indicator of target word identity.

In addition, children's ability to integrate top-down information into bottom-up processing may be contingent upon the amount of time that children have to process and deploy the available top-down information before the relevant bottom-up input arrives. Let's consider the two forms of top-down cues that are available in studies of prepositional phrase attachment ambiguity: referential context cues and modifier plausibility cues. In a sentence such as *Tickle the frog with the fan*, top-down inferences about reference are initiated by the mention of the word *frog*, which appears only one word prior to the onset of the *with*-phrase. Constructing inferences about referential likelihood based on the visual context and the speaker's referential model take time and effort to compute (see Altmann & Steedman, 1988; Yacovone et al., 2021 for discussion of the multi-step feedback loops required to generate and incorporate referential inferences into sentence comprehension). If children do not have the time to construct and make top-down inferences based on the referential context by the onset of the prepositional phrase (due to less efficient processing speed; e.g., Hale, 1990; Kail, 1991; Kail & Salthouse, 1994), they may rely on readily accessible intonational cues and verb information to make an initial parse. Top-down inferences about modifier plausibility are not available until the final word of the *with*-phrase (*fan*), which arrives after this initial parse is made. As discussed above, it may be difficult for young children to override this initial parse if the plausibility information contradicts

their initial interpretation due to reduced age-related differences in inhibitory control (Carver et al., 2001; Johnstone et al., 2005; Passler et al., 1985; Welsh et al., 1991; Wiebe et al., 2012; inter alia). In contrast, in the present study, the top-down information cueing the target word was typically available well before target onset (on average four words prior to the target; see *§3.4.1.1. Incremental cloze task*), providing ample time for this information to propagate through the language comprehension system before the arrival of the target word in the bottom-up input.

Finally, if we limit ourselves to spoken language comprehension, children appear to succeed in using top-down information to guide lexical processes (e.g., Hahn et al., 2015; Levari & Snedeker, 2024) but fail in using top-down constraints to guide syntactic processing (e.g., Kidd et al., 2011; Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008; Trueswell et al., 1999; Yacovone et al., 2021). This raises the possibility that there are broad differences between syntactic and lexical processes that make integrating top-down cues into syntactic processing more difficult and costly for young children. While representations of words are stored in the mental lexicon, syntactic representations cannot be fully stored and must be constructed on the fly, at least some of the time (allowing for productivity). Generating candidate syntactic representations and assessing their fit to the top-down context may require more cognitive resources than activating and evaluating stored lexical representations, making this process more challenging for young children to execute during real-time sentence processing. Furthermore, the utility of prediction may be lower for syntactic processing than lexical processing, resulting in a poorer cost to benefit ratio: Whereas word recognition requires speakers to identify the correct lexical item from the tens of thousands of candidates stored in the mental lexicon, the possible candidate representations for real-time syntactic parsing are far fewer — for instance, when a listener encounters the word *with* in the sentence *Tickle the frog with the fan*, there are two

primary syntactic constructions for them to select between (attaching the preposition to the noun *frog* or the verb *tickle*). When there are more candidates to discriminate during processing, it may be more worthwhile to delegate cognitive resources to pre-activating possible representations and integrating top-down information into the evaluation of candidates. Therefore, the contrast between prior work and the present study may reflect the relative feasibility and benefits of top-down integration in syntactic as opposed to lexical processing.

In sum, the present study taken together with the body of work on syntactic ambiguity resolution suggests that although the pathways for top-down comprehension are in place from early childhood, information doesn't always travel through them rapidly enough for real-time integration. Children's ability to integrate top-down information into language comprehension may depend on factors such as the form of processing (lexical vs. syntactic), the reliability and accessibility of the available top-down cues, and the time children have to integrate these cues into processing (see Yacovone et al., 2021). Determining the conditions under which top-down contextual cues are integrated into young children's language comprehension is an important avenue for future research, with implications for theories of language processing and acquisition, how language interacts with other cognitive functions such as attention, memory, and executive functioning, as well as for education and language interventions.

3.5.3. The role of top-down information in children's language development

The present findings illustrate that top-down pathways are in place early in childhood, prior to formal schooling and literacy. This early availability of top-down information during language comprehension may play an important role in young children's language acquisition and literacy development.

Access to top-down information and the ability to predict upcoming referents and lexical representations has the potential to assist young children in learning new words (see Rabagliati et al., 2016 for review). Indeed, many studies have observed correlations between young children's vocabulary sizes and referential prediction abilities (e.g., Borovsky et al., 2012; Lew-Williams & Fernald, 2007; Mani & Huettig, 2012), though these correlations on their own do not reveal whether prediction improves with vocabulary expansion or drives it. There are several means by which top-down processing may facilitate novel word learning. First, in cases when contextual information allows a child to predict an upcoming referent for which they do not have a label, having anticipated this message-level information could help the child to map the word form to the referent (facilitating the search for its meaning) and could help them avoid activating words that are temporarily consistent with the bottom up input and thus might interfere with learning the correct meaning (e.g., *candy* or *candle* when first encountering *candelabra*). Second, in cases when a child uses the available context to predict an upcoming lexical representation, if the child then encounters a novel word in its place, this might allow the child to make inferences about the relationship between the novel word and the known word — i.e., that the novel word appears in similar contexts to a known familiar word and may share semantic features (see Lany & Saffran, 2010 for evidence that infants use distributional information to learn novel words). In fact, Reuter et al. (2019) found that when a novel word appeared in a context in which a familiar referent was expected, three to five year-olds' success in learning the novel word was in part predicted by whether they anticipated the familiar referent (and perhaps also the novel word). Third, if a child correctly predicts upcoming word representations and uses top-down information to constrain bottom-up processing, this will facilitate processing of familiar words, which may give the child more bandwidth to attend to and learn from any novel information

appearing later in the input (Fernald et al., 2008; Fernald et al., 2006). Novel word learning may be facilitated by any or all of these processes. An important direction for future research is tracing when top-down interactivity first comes online for infants and how it relates to the trajectory of lexical acquisition — for instance, testing whether the rapid increase in children's word comprehension in the second year of life (Bergelson & Swingley, 2012; Bergelson & Swingley, 2013; Bergelson & Swingley, 2015) could result in part from increases in the accessibility of top-down information (e.g., Bergelson, 2020).

Beyond word learning, the availability of top-down information during comprehension may influence language development as children learn how to read. Learning to read provides children with a new form of language input containing diverse vocabulary and syntactic structures (e.g., Cunningham & Stanovich, 1998; Nation et al., 2022), expanding children's opportunities for linguistic growth. Children who are learning to read alphabetic languages acquire two kinds of mappings: (i) grapheme to phoneme correspondences and (ii) lexical links between existing phonological and semantic representations of words and new orthographic representations (Castles et al., 2018). Learning both of these mappings could be facilitated by top-down prediction. When children are first learning to read, the bottom up decoding of the input slows down to a crawl as they carefully sound out each letter, leaving ample time for top-down constraints to influence lower level processes. If top-down information picks out the correct word before decoding begins, activation of the word's phonological form, coupled with the visual input, could reinforce weak grapheme–phoneme correspondences that the child might otherwise forget or may weed out competing correspondences that are locally consistent with the input (for example, if you are reading about a child flying, and you see "pla…", the *a* is probably an /eɪ/ as in *plane* not an /æ/ as in *plan*). As reading progresses, the strategy of sounding out a

word is replaced with whole-word retrieval (Castles et al., 2018). Building up these lexical mappings could be facilitated by top-down prediction allowing the full phonological form of the word, and its meaning, to be available when the word is fixated (providing a mapping that does not depend on strictly-ordered decomposition). Books for early readers are notoriously repetitive and predictable, suggesting that our educational practices (and our naive theory of the child) assume that lexical prediction emerges before literacy.

The relationship between literacy and prediction may also be bi-directional. Acquiring orthographic representations may influence the accessibility of phonological forms for pre-activation (literacy restructures phonological representations and increases phonological awareness; e.g., Morais et al., 1979; Pattamadilok et al., 2010; Perre et al., 2009) and sharpen existing lexical representations (Mani & Huettig, 2014), further enhancing young children's lexical prediction skills. Indeed, prior work has posited that learning to read improves predictive processing in spoken language (e.g., Huettig & Pickering, 2019), though the evidence for this relationship is primarily derived from studies with adults (e.g., Huettig & Brouwer, 2015; James et al., 2023; Mishra et al., 2012; Ng et al., 2018; cf. Mani & Huettig, 2014) and using measures of referential prediction rather than lexical prediction (cf. Ng et al., 2018). To test these hypotheses about the role of literacy in prediction (and prediction in literacy), we must study children as they first encounter print. Our work sets the stage for these investigations by demonstrating how form-based prediction can be studied in young children.

3.6. Conclusion

Cognition involves interactive processing as top-down and bottom-up pathways pass information between different levels of representation. The present work provides insight into

the development of these mechanisms, using language as a case study. We found evidence that these pathways interact during language comprehension by the fourth and fifth years of life: The children in our study were able to use top-down cues about sentence context to constrain the bottom-up processing of the linguistic input as well as to pre-activate, or predict, word's phonological form representations before they appeared in the input. The availability of top-down information during language comprehension in early childhood may play a key role in children's language acquisition and literacy development. More generally, these results suggest that the pathways required for interactive processing are available early, which could indicate that they are a fundamental consequence of the mind's architecture. This interpretation of the results predicts that interactive processing should arise early in other domains (e.g., face and object recognition) as well as in the other languages both spoken and signed, suggesting new avenues for developmental research.

Chapter 4

[Paper 3]

Assessing two methods of webcam-based eye-tracking for child language research

Margaret Kandel & Jesse Snedeker

Published 2024 in Journal of Child Language

[https://doi.org/10.1017/S0305000924000175]

4.1. Introduction

Visual-world eye-tracking is an important tool for studying real-time language processing in children. In the visual-world paradigm, participants are presented with a display, and their eye-movements are recorded as they listen to or produce an utterance. Individuals systematically look to referents or associates of the words they hear (e.g., Cooper, 1974; Tanenhaus et al., 1995) or are planning to produce (e.g., Griffin & Bock, 2000; Meyer et al., 1998). Saccades are tightly linked to linguistic information, with fixations to relevant stimuli rising within 200 ms of the onset of linguistic cues in adults (e.g., Allopenna et al., 1998; Cooper, 1974). This relationship has allowed researchers to use eye-movements to investigate a variety of questions in language processing (see Huettig et al., 2011 for review). This paradigm is particularly useful for child research, as it provides a non-invasive, real-time measure of language processing that doesn't require meta-linguistic reasoning (cf. grammaticality judgments, lexical decision), reading ability (cf. self-paced reading), or a lengthy set-up (cf. electroencephalography). Children similarly look

to relevant stimuli shortly after the onset of linguistic cues, and visual-world experiments have been used with children to study multiple levels of language processing, including phonological (e.g., McMurray et al., 2018; Sekerina & Brooks, 2007), morphological (e.g., Özge et al., 2022; Zhou et al., 2014), syntactic (e.g., Contemori et al., 2018; Snedeker & Trueswell, 2004; Trueswell et al., 1999), semantic (e.g., Borovsky et al., 2012; Brouwer et al., 2019), and pragmatic processing (e.g., Cooper-Cunningham et al., 2020; Huang & Snedeker, 2009; Kampa & Papafragou, 2020).

Visual-world experiments are primarily conducted in university labs where researchers employ specialized equipment to monitor participant gaze (e.g., SR Research, 2021; Tobii, 2021). More recently, however, algorithms that determine gaze location based on webcam video have increased interest in conducting eye-tracking experiments without specialized equipment and outside of lab settings (e.g., Erel et al., 2022; Fraser et al., 2021; Papoutsaki et al., 2016; Valenti et al., 2009; Valliappan et al., 2020; Xu et al., 2015). Webcam-based eye-tracking allows researchers to conduct experiments over the internet, in either supervised settings (with an experimenter present over video conferencing) or unsupervised settings (with no experimenter present). Web-based testing has several advantages, many of which are particularly relevant to child research. Participants can complete experiments from the comfort of their own homes, where children may feel more at ease. This frees families from needing to travel to the lab and make babysitting arrangements for siblings. Unsupervised web-based experiments allow for even more efficient data collection, as sessions can occur outside of working hours at whatever time is most convenient for families. Collecting data over the internet gives researchers access to more diverse populations (see Henrich et al., 2010 for the importance of sample diversity) and languages not spoken near their home institutions. Webcam-based eye-tracking can also be used

in conjunction with direct participant contact, allowing researchers to set up mobile labs wherever they can bring a laptop (e.g., schools, parks, museums, etc.).

Of the algorithms that track eye-gaze from webcam videos, the JavaScript library *WebGazer.js* (hereafter "WebGazer"; Papoutsaki et al., 2016) has garnered the most attention from behavioral researchers. WebGazer is open-source and has been integrated into popular frameworks for running online behavioral tasks, such as PCIbex (Zehr & Schwarz, 2018), JsPsych (de Leeuw, 2015), and Gorilla (Anwyl-Irvine et al., 2020). Gaze estimation occurs locally in the user's web-browser, and no video is saved, thus maintaining participant privacy. Although initially designed to detect eye-gaze during user interactions with webpages (Papoutsaki et al., 2016), recent studies have explored WebGazer's suitability for behavioral research with adults.

The results of these investigations are promising. WebGazer detects looks to perceptual stimuli shortly after they appear (e.g., Semmelmann & Weigelt, 2018; Slim & Hartsuiker, 2022) and has been used to replicate previously-observed eye-tracking effects in a variety of domains, including visual inspection of faces (Semmelmann & Weigelt, 2018), decision making (Yang & Krajbich, 2021), and language processing (Degen et al., 2021; Slim & Hartsuiker, 2022; Vos et al., 2022). However, WebGazer has limitations compared to the eye-tracking devices typically used for in-lab studies. Specifically, the offset between estimated gaze and stimulus locations is greater and looking patterns are delayed relative to in-lab studies (e.g., Degen et al., 2021; Semmelmann & Weigelt, 2018; Slim & Hartsuiker, 2022). At present, it is not clear to what extent this noise is attributable to WebGazer itself as opposed to properties of the less controlled web-based setting (e.g., variations in software, hardware, environments, and internet connections) or differences in participant behavior when completing studies online.

Given these findings with adults, it seems reasonable to consider using WebGazer for web-based psycholinguistic studies with children. However, it is not obvious that WebGazer would perform as well when estimating child gaze. Child faces are smaller than those of adults, and children are likely to be in a different position relative to the webcam because of their height, which could reduce the accuracy of WebGazer's pupil detection and gaze estimation algorithms. In addition, young children are less likely to remain in the same position for the duration of a task, and they are unlikely to have the patience to sit through extensive calibration/recalibration procedures that improve accuracy in adult studies (e.g., Semmelmann & Weigelt, 2018; Yang & Krajbich, 2021). In fact, even high-end in-lab eye-trackers are less accurate when used with children (Dalrymple et al., 2018). Furthermore, children may have more difficulty maintaining attention when completing an experiment from home, where there may be more distractions than in controlled lab settings.

In the present study, we investigate whether it is possible to run web-based visual-world studies with school-aged children. We test two webcam eye-tracking methods: automatic gaze estimation with WebGazer and frame-by-frame annotation of gaze direction (e.g., Snedeker & Trueswell, 2004) from webcam videos recorded via Zoom teleconferencing software (https://zoom.us/). Experiment 1 directly compares these two methods in a visual-world language task with five to six-year-old children. We assess how well these methods discriminate both robust fixation patterns (looks to target stimuli) as well as more subtle eye-movement patterns of the kind relevant to child language researchers (phonemic cohort competition effects; e.g., Allopenna et al., 1998; Sekerina & Brooks, 2007). By collecting both forms of gaze data simultaneously, we can assess the extent to which any noise observed in the WebGazer data stems from WebGazer itself as opposed to participant behavior or the web-based setting.

Experiment 2 focuses more specifically on WebGazer, assessing its performance with child participants aged four to twelve years in a visual-fixation task. Experiment 2 was run without an experimenter present, allowing us to assess the feasibility of conducting unsupervised web-based eye-tracking studies with child participants.

4.2. Experiment 1: Visual-world task

Experiment 1 comprised two linked experiments focused on the phonemic cohort competition effect. This effect is well-suited for testing the efficacy of web-based visual-world eye-tracking, as it has been replicated many times with both adults (e.g., Allopenna et al., 1998; Dahan & Gaskell, 2007; Dahan et al., 2001; Farris-Trimble & McMurray, 2013; Magnuson et al., 1999; inter alia) and children (e.g., Desroches et al., 2006; Sekerina & Brooks, 2007; Rigler et al., 2015; Weighall et al., 2017; inter alia), and the presence of cohort activation is often used to investigate higher-level linguistic constraints on incremental language processing (e.g., Dahan & Tanenhaus, 2004; Gaston et al., 2020; Ito et al., 2018; Li et al., 2022; Paul et al., 2019). In a visual-world context, cohort competition effects arise when listeners hear a target word that shares onset phonemes with one of the images on the screen; when hearing the onset of the target word (e.g., beaker), listeners fixate more on the image of a cohort competitor (e.g., beetle) than phonologically-unrelated distractors (e.g., carriage) (e.g., Allopenna et al., 1998). The onset of competition effects follows a similar time-course in both adults and children, though effects continue longer in young children (Sekerina & Brooks, 2007).

Experiment 1 used two different visual displays to see how each is affected by the noise introduced in web-based experimentation. Experiment 1A used a simple two-image display (with images on the left and right), similar to many infant preferential-looking studies. Experiment 1B

used the four-image display that is common in visual-world studies (one image in each quadrant). Experiment 1B's four-image display further allows us to assess the performance of the eye-tracking methods on horizontal and vertical look discrimination.

The experiment methods and WebGazer phonemic cohort analysis were preregistered (https://osf.io/cn3ur/). The analysis of the webcam video data was exploratory. Prior to conducting Experiment 1, we ran a pilot experiment (N=24) to assess WebGazer's performance with adult participants (see Supplementary Materials; https://osf.io/hmeyb/).

4.2.1. Methods

A more detailed description of the methods is available in the Supplementary Materials. All experiments reported in this paper were approved by the Harvard University-Area Committee on the Use of Human Subjects. Data, analysis code, and Supplementary Materials are available from https://osf.io/hmeyb/.

4.2.1.1. Participants

Experiment 1 had 64 participants of five and six years of age who were native monolingual speakers of American English. Half completed Experiment 1A (N = 32, 14 F, 18 M; M_{age} = 5.8 years, SD = 0.6, range = 5;0–6;11), and half completed Experiment 1B (N = 32, 20 F, 12 M; M_{age} = 6.2 years, SD = 0.5, range = 5;0–6;11). Our sample size (32 participants per experiment) is similar to psycholinguistic experiments in general and to previous studies of the phonemic cohort effect (e.g., Farris-Trimble & McMurray, 2013; Huettig & McQueen, 2007). Informed written consent was received from the parent or guardian for their child's participation. Participants were compensated with a $5.00 gift card.

4.2.1.2. Materials

We selected 36 target–cohort pairs with onset overlap of one or more phonemes. As a control, each target word was pseudo-randomly assigned a competitor from another target–cohort pair with no onset overlap. The experiments consisted of 36 trials (one per word pair). The trial displays included a target image (corresponding to the target word) and a competitor image. In Experiment 1B, the displays also included two pseudo-randomly assigned distractor images whose names had different onsets from the target and competitor.[1] The trials were rotated through two conditions in two presentation lists. In the cohort condition, the competitor image depicted the cohort pair of the target (e.g., the target *milk* appeared with the competitor *mitten*). In the control condition, the target appeared with its control competitor (e.g., the target *milk* appeared with the competitor *windmill* from the cohort pair *window–windmill*). The cohort effect was assessed by comparing looks to the competitor images in the cohort and control conditions.

The experiments were built in PCIbex (Zehr & Schwarz, 2018) using PCIbex's implementation of WebGazer v2 and were completed in the participant's web-browser. To accommodate the variability in screen-sizes across participant computers, stimulus size and location were defined by browser window size (equivalent to screen-size since the experiment was displayed fullscreen). Images appeared on canvases centered in their quadrant or half of the screen (Figure 4.1). Throughout each trial, WebGazer tracked looks to these canvases. When WebGazer detected a look to a canvas, the canvas border turned purple.[2]

[1] One target (*doctor*) was accidentally assigned a distractor (*dolphin*) that shared onset sounds, so this trial was omitted from the Experiment 1B analysis.

[2] This color-change functionality allowed participants to use their eyes to select images from the screen (see Supplementary Materials for additional information about the task instructions given to participants). Initial piloting of web-based tasks with young children revealed that they were not familiar with how to

Figure 4.1: *Example Experiment 1A (left) and Experiment 1B (right) trials.* Each competitor image (e.g., *mitten*) appeared with its own target in the cohort condition (e.g., *milk*, right) and with another target in the control condition (e.g., *banana*, left). Image canvas borders turned from gray to purple when WebGazer estimated eye-gaze to fall on the image. Stills include images from Duñabeitia et al. (2018) and Rossion and Pourtois (2004).

4.2.1.3. Procedure

Participants completed the experiment while in a Zoom teleconference call with the experimenter(s), and the session was recorded via the Zoom meeting recording function. The participant opened the link to the experiment on their computer in Google Chrome or Mozilla Firefox and used the Zoom screen-sharing function to share the display with the experimenter. Participants using a non-Mac computer (with the exception of one Chromebook user) turned off their Zoom video prior to opening the experiment, as piloting revealed that many of these computers do not allow the same webcam to be used by Zoom and WebGazer simultaneously.

use a computer mouse or trackpad, and click-based selection responses thus prompted a large number of participant looks directed at these tools instead of on the screen. Piloting with the color-change functionality indicated that it kept participants' attention on the screen, gave them a sense of agency in the task, and was not distracting.

At the beginning of the experiment, the participant completed an audio check and a WebGazer calibration sequence. As we were interested in the range of calibration accuracy that would be obtained with our sample, we did not specify a minimum calibration threshold. After calibration, participants completed three practice trials followed by the 36 experimental trials. Each trial started with a calibration check. Next, the images appeared. After 2000 ms, participants heard pre-recorded audio instructions telling them to *Look at the + [target word]*. The images remained on screen for 2250 ms after audio offset. The full experiment session took approximately 20–30 minutes.

4.2.2. Analysis

The data for Experiments 1A and 1B were analyzed separately. All analyses were conducted using R v4.1.0 (R Core Team, 2021).

4.2.2.1. WebGazer

In each trial, WebGazer recorded looks from trial onset to two seconds after audio offset. In each sample, a 0 or 1 was recorded for each image canvas indicating whether or not participant gaze fell upon it (0 = no, 1 = yes). Sampling rate varied by participant, likely dependent upon their computer, webcam, and internet connection (grand mean time between samples = 96ms, SD = 43ms).[3] Samples which recorded no looks to any of the image canvases

[3] Note that this average sampling rate (approximately 10 Hz) is slower than observed in some other WebGazer investigations, in which sampling rates range from 14–21 Hz (e.g., Prystauka et al., 2023; Semmelmann & Weigelt, 2018; Vos et al., 2022). Vos et al. (2022) and Prystauka et al. (2023) both implemented exclusion criteria to omit participants with a sampling rate below 5 Hz. Applying this same exclusion criteria to Experiment 1 (resulting in omission of $n = 2$ participants), the mean time between samples is 93 ms (SD = 38ms), or approximately 11 Hz, suggesting that the slower sampling rate observed in Experiment 1 is not due to the lack of exclusion criteria but rather reflects the variability of web-based experimentation.

were excluded from analysis (41.24% of Experiment 1A samples; 27.07% of Experiment 1B samples). To regularize sampling rates prior to analysis, we analyzed gaze locations in bins of 100 ms. A time bin received a value of 1 for a canvas if at least 50% of recorded looks within the bin fell on that canvas.

We preregistered a cluster permutation analysis to investigate competitor looks 0–2000 ms after target onset (e.g., Hahn et al., 2015; Yacovone et al., 2021). This analysis assessed the effect of interest at each time step using generalized linear mixed-effect models (GLMMs) with a binomial distribution and logit link (step size = 100 ms).[4] All models in the present study were fit using the {lme4} package v1.1-27.1 (Bates et al., 2015). The models had looks to the competitor image (0, 1) as the dependent variable, a fixed effect of condition (cohort, control), and random slopes and intercepts for condition by participant and item. Item was individuated by competitor image identity to account for variance in properties of the competitor images. An effect was considered reliable at a step if the absolute value of its *z*-value was greater than 2 (Gelman & Hill, 2007).[5] A minimum of two sequential reliable effects were required to comprise a cluster. To assess cluster reliability, we performed 1000 simulations reshuffling the condition labels for

[4] It is important to note that while cluster-based permutation analyses provide information about the presence of effects, they cannot be used to make inferences about the onset and duration of these effects (for discussion, see Fields & Kuperberg, 2019; Groppe et al., 2011; Sassenhagen & Draschkow, 2019). As there are no corrections for multiplicity, false positives may emerge in the initial cluster identification, meaning that researchers cannot make inferences about effect significance at any one time bin in the cluster (including the first or final time bins). In addition, the cluster-mass permutation test does not assess how adding or removing time bins from the cluster (e.g., at the beginning or end) influences its overall reliability. Furthermore, cluster duration is sensitive to data quantity, power, and the chosen threshold for including time bins within a cluster, which could lead to under- or overestimations of the extent of effects.

[5] If a model failed to converge at a step (excluding singular fit warnings), we did not use the computed model estimates for that step. Instead, following Yacovone et al. (2021), we used the model estimates from the prior step; if the model at the first step did not converge, the *z*-value was set to zero. This procedure prevents models that do not converge properly from breaking up or prematurely ending a cluster. There were no steps with non-convergence in the analyses of the observed data.

each participant. In each simulation, we summed the *z*-values of the adjacent steps in identified clusters to obtain a *z*-sum statistic. We compared the *z*-sum of the observed cluster to the distribution of each simulation's largest *z*-sum. A *p*-value for the observed cluster was determined by its position in this distribution (e.g., for a *p*-value of <0.05, 95% of the *z*-sums in the distribution must be greater than or equal to the observed statistic).[6]

We also analyzed the effect of condition on competitor looks in two time windows: 300–700 ms after target onset (preregistered) and 600–1000 ms after target onset (exploratory to account for a potential WebGazer delay in look detection). The results of these analyses are broadly consistent with the findings from the cluster analyses reported below and appear in the Supplementary Materials.

We conducted an additional exploratory analysis to investigate when target image looks were reliably different from chance in each condition. For each condition, we performed cluster permutation analyses assessing looks to the side of the screen containing the target image 0–2000 ms after target onset; for Experiment 1B, we performed separate analyses for the horizontal and vertical side distinctions. In Experiment 1A, a look was considered to fall on the same side of the screen as the target if it fell on the target image; the analysis thus assesses the likelihood of target image looks. In Experiment 1B, a look was considered to fall on the same side of the screen as the target if it fell on the target or on the image vertically-adjacent (for the horizontal-side analysis) or horizontally-adjacent (for the vertical-side analysis). The analyses followed the same procedure described above, except that to assess reliability, we reshuffled the trial image location configurations by participant (thus preserving for each participant the overall number of target

[6] In this analysis, it is possible to produce a *p*-value equal to zero if 0% of *z*-sums in the distribution of simulated statistics are greater than or equal to the observed statistic. We report these *p*-values as $p < 0.001$.

and non-target images appearing in each quadrant). The GLMMs computed at each step had target side looks (0, 1) as the dependent variable and random intercepts for participant and item (i.e., target image identity); as the model had no fixed effect, the likelihood of target side looks was compared to chance (50%). This analysis allows us to identify when each method is able to discriminate looks to the target quadrant along both the horizontal and vertical dimensions. For Experiment 1B, we supported the results of this analysis with a multinomial regression analysis assessing when looks differed between the target and the horizontally-, vertically-, and diagonally-adjacent images (see Supplementary Materials); the results align with the target side looks analyses.

4.2.2.2. Webcam video annotation

To gain further information about the eye-gaze patterns of our participants, we hand annotated gaze direction in the webcam videos of all participants who were able to keep their Zoom video on as they completed the experiment. Trial onsets times were identified from Zoom screen recordings using Python scripts that detected when the colored stimulus images appeared on screen (Anthony Yacovone, personal communication). These onsets were used to divide the continuous webcam videos into separate trial videos. Coders (blind to condition and target/competitor location) annotated gaze direction for each frame of these videos (annotation script by Anthony Yacovone).

Paralleling the WebGazer analysis, samples that were not coded as looks to one of the image locations were removed from analysis (i.e., center looks, blinks, etc.) (34.72% of Experiment 1A samples; 23.24% of Experiment 1B samples). The webcam videos had 40 ms

between samples. To compare to the WebGazer data, we analyzed gaze locations in bins of 100 ms, following the binning procedure described above.

All videos were annotated by a single coder. To assess reliability, each video was additionally annotated by a secondary coder. Within our cluster analysis window (0–2000 ms after target onset), inter-coder agreement was 92.18% in the Experiment 1A dataset and 90.02% in the Experiment 1B dataset (see Supplementary Materials for details). We performed the same analyses on the webcam video data as on the WebGazer data.

4.2.3. WebGazer results

Ten Experiment 1A trials across eight participants and 18 Experiment 1B trials across seven participants were omitted from the WebGazer analysis because no data were saved for them on our server.

4.2.3.1. Calibration scores

Participant calibration scores in the initial calibration sequence ranged from 2–80% across Experiments 1A and 1B, with an average of 43% (SD = 18, see Supplementary Materials for plots and more detail). Mean participant calibration scores during the calibration checks at the beginning of each experimental trial ranged from 8–50%, with an average of 30% (SD = 11).

4.2.3.2. Experiment 1A

Figure 4.2 illustrates the increase in looks to the target image in the WebGazer output following target word articulation in both the cohort and control conditions. This pattern was similar for targets on the left and right of the screen (see Supplementary Materials). While there

was a substantial rise in target looks in both conditions (~75% of looks), this rise was smaller than commonly observed in two-image studies with children and adults (e.g., 80–85% with adults and three to four-year-olds in Simmons, 2017).

Target looks were reliably different from chance in clusters starting 800 ms after target onset in the control condition (z-sum = 64.94, $p < 0.001$) and 1000 ms after target onset in the cohort condition (z-sum = 61.82, $p < 0.001$).

Figure 4.3 focuses on the cohort effect by plotting looks to the competitor image in the cohort and control conditions. Prior to target word onset, looks to the competitor image were at chance (50%). These looks began to decline approximately 700 ms after target word offset (as target looks increased). Our analyses explored whether this decline was faster in the control condition than the cohort condition. The analysis identified a reliable difference in competitor looks between conditions in a cluster 900–1099 ms after target onset (z-sum = 4.87, $p = 0.02$).

Figure 4.1: *Mean WebGazer looks to the target and competitor images by condition in Experiment 1A.* Ribbons indicate standard error. Vertical lines indicate average target word duration. Shading indicates when looks to the target image differed from chance.

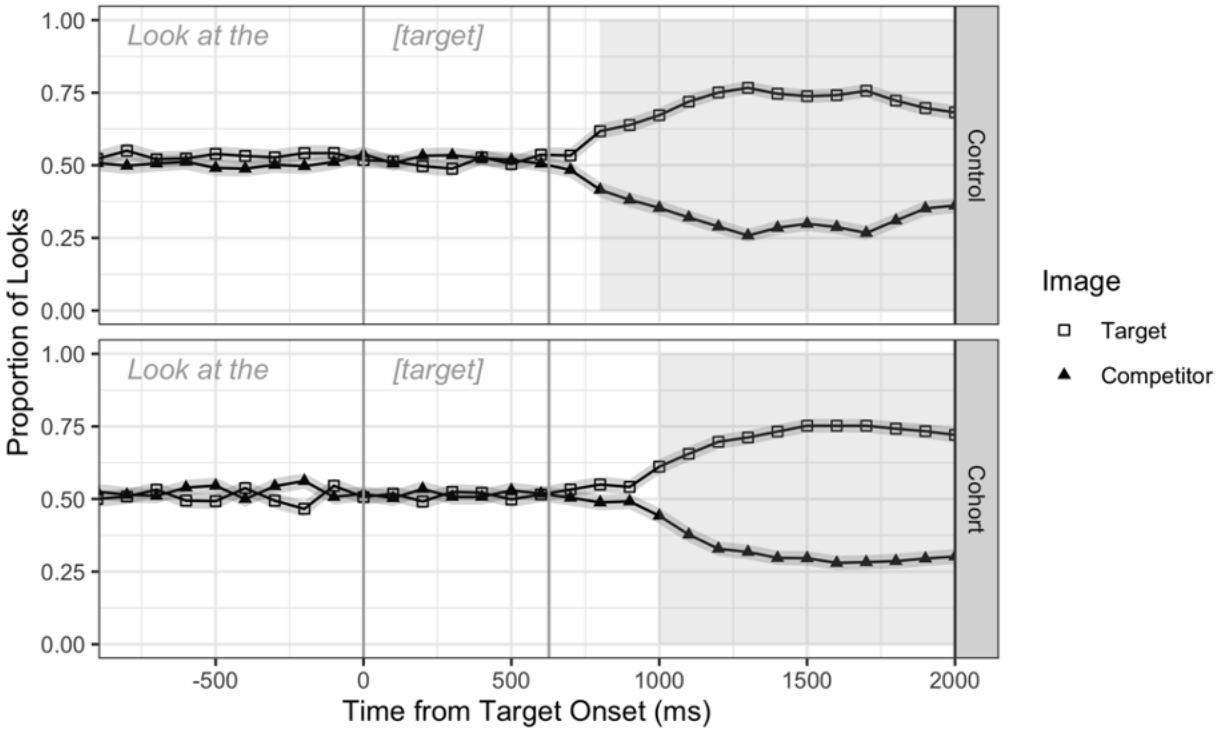
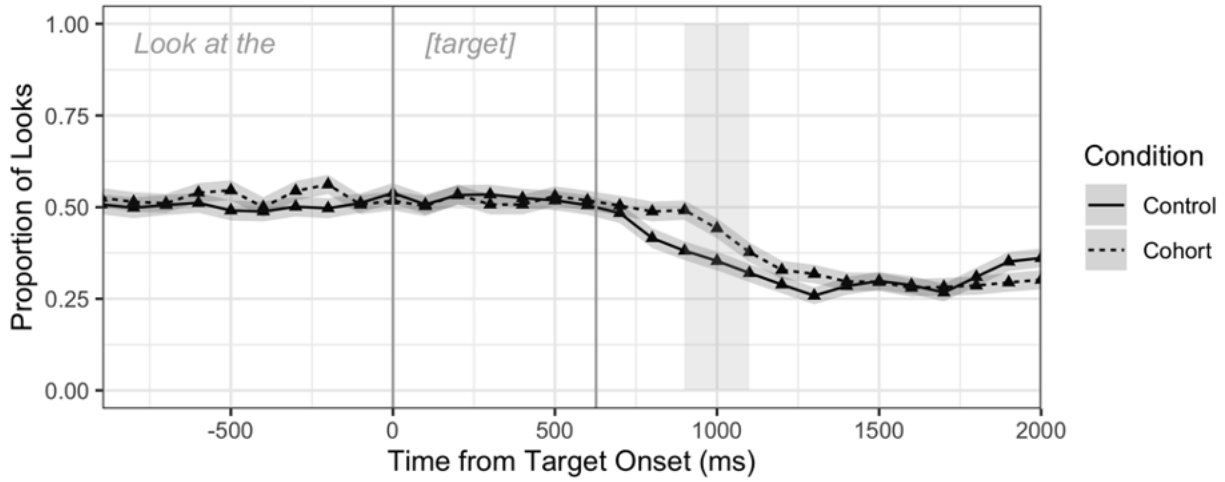


Figure 4.3: *Mean WebGazer looks to the competitor image by condition in Experiment 1A.* Ribbons indicate standard error. Vertical lines indicate average target word duration. Shading indicates when looks between conditions were reliably different in the cluster analysis.



4.2.3.3. Experiment 1B

Figure 4.4 shows looks to the target image, competitor image, and two distractor images (collapsed) as detected by WebGazer in the cohort and control conditions. In both conditions, WebGazer detected increased looks to the target image following target word onset. However, the effects appeared smaller than in previous studies (≤ 50% in the present study vs. > 60% with five and six-year-olds in Sekerina & Brooks, 2007).

In the control condition, looks to the side of the screen containing the target were reliably different from chance in clusters starting 900 ms after target onset along the horizontal axis (z-sum = 62.73, $p < 0.001$) and 1200 ms after target onset along the vertical axis (z-sum = 29.10, $p < 0.001$). In the cohort condition, clusters emerged 1000 ms after target onset for the horizontal-side distinction (z-sum = 50.49, $p < 0.001$) and 1400 ms after target onset for the vertical-side distinction (z-sum = 17.83, $p = 0.001$).

The observed clusters for the horizontal-side distinction had similar onsets to those in Experiment 1A (800 ms in the control condition, 1000 ms in the cohort condition) — however, the observed clusters for the vertical-side distinction started 300–400 ms later, suggesting that WebGazer may have more difficulty discriminating looks along the vertical axis. Figure 4.5 plots participant looks to the target and distractor (non-target) images in the control condition 1200–2000 ms after target onset (when participants were likely fixating on the target quadrant, according to WebGazer). In this window, there were more looks to the vertical distractor than the other non-target images, supporting the hypothesis that WebGazer has increased difficulty discriminating vertical looks (this pattern was confirmed in an exploratory multinomial analysis; see Supplementary Materials). A figure showing target and distractor looks by target location is available in the Supplementary Materials.

Figure 6 plots looks to the competitor image in the cohort and control conditions. Prior to target word onset, looks to the competitor image were at chance (25%). These looks began to decrease approximately 1200 ms after target onset. The cluster analysis did not identify any clusters where competitor looks differed in the two conditions. Thus, we did not replicate the phonemic cohort effect.

Figure 4.4: *Mean WebGazer looks to the target image, competitor image, and distractor images (collapsed) by condition in Experiment 1B.* Ribbons indicate standard error. Vertical lines indicate average target word duration. Shading indicates the temporal overlap of the clusters when target side looks differed from chance in both the horizontal and vertical directions.

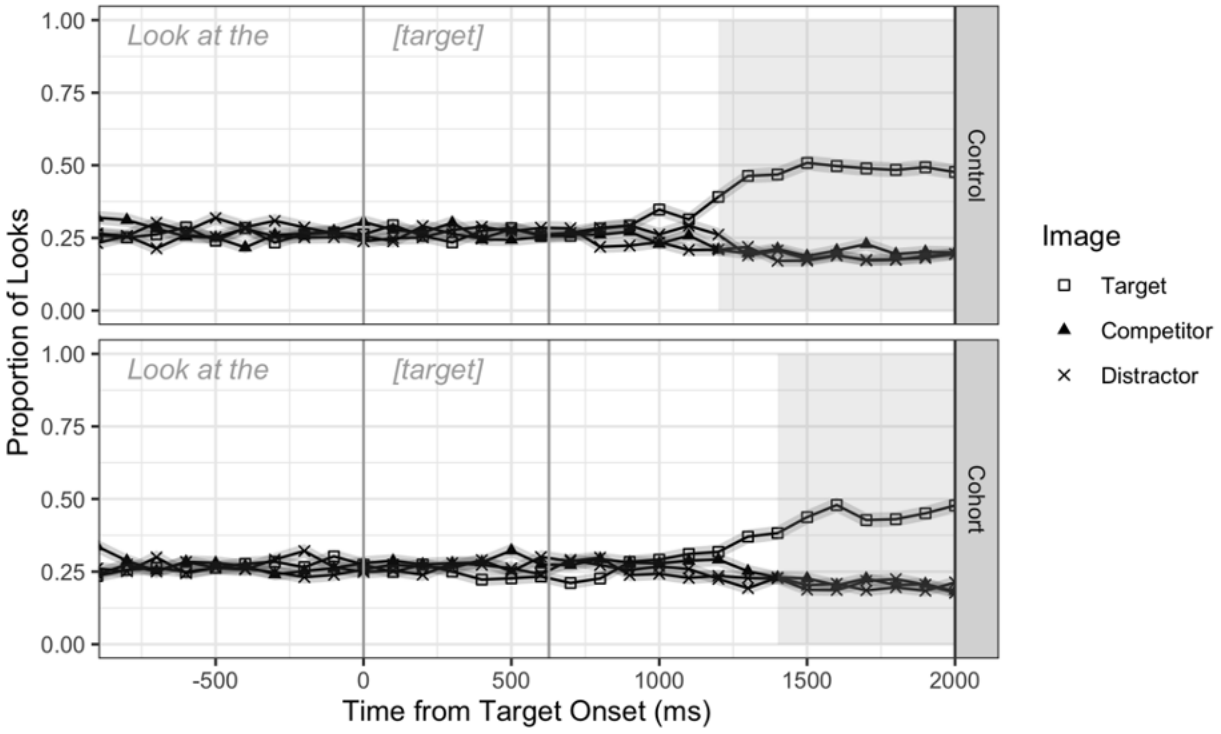


Figure 4.5: *Boxplot of participant WebGazer fixation proportions to the target and non-target images in the Experiment 1B control trials from 1200–2000 ms after target onset.* Mean fixation proportions for each image are labeled and identified by black diamonds. The gray points represent participant means.

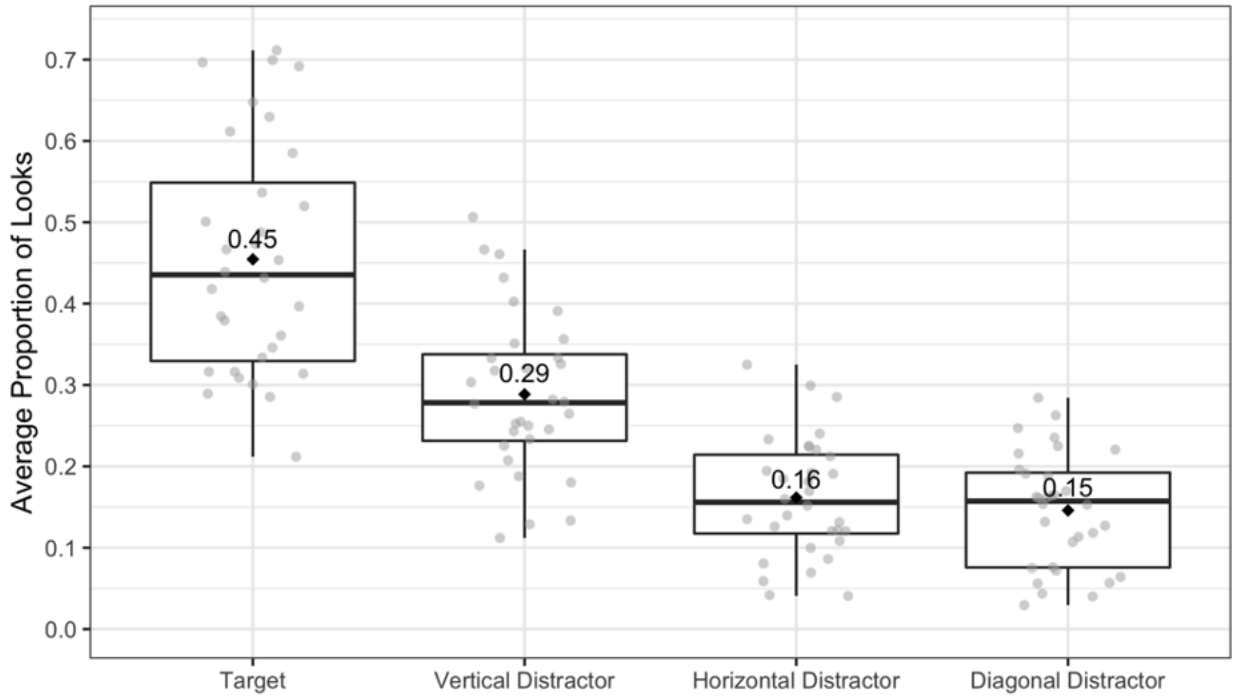
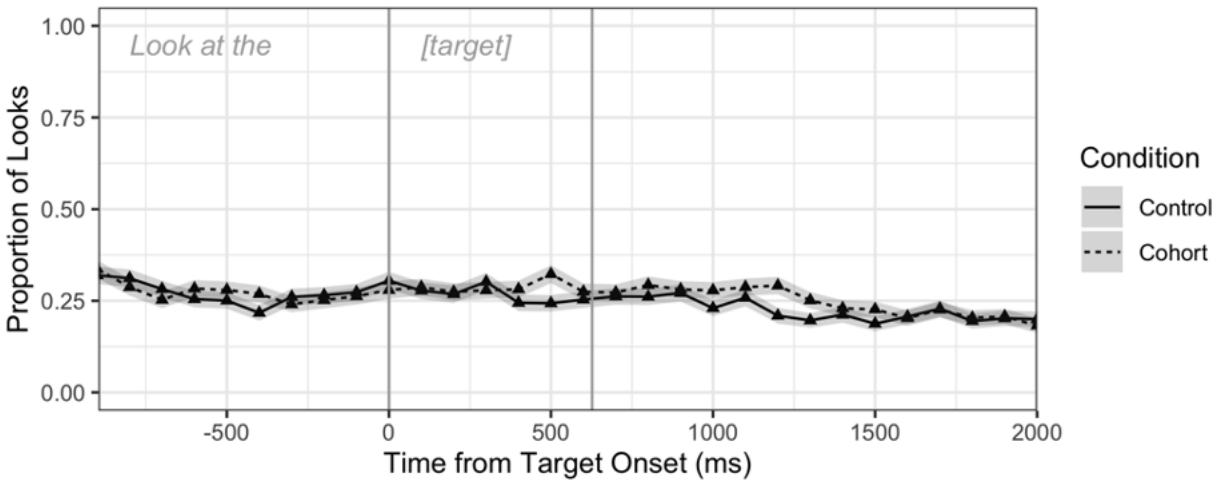


Figure 4.6: *Mean WebGazer looks to the competitor image by condition in Experiment 1B.* Ribbons indicate standard error. Vertical lines indicate average target word duration.



4.2.4. Webcam video annotation results

We had video data for 13 of 32 participants for each experiment.

4.2.4.1. Experiment 1A

Figure 4.7 plots looks to the target and competitor images in the cohort and control conditions as detected by hand annotation and WebGazer for the 13 participants with video data. Webcam video annotation identified a higher proportion of target image looks than WebGazer. The pattern of performance was similar for targets on the left and right of the screen (see Supplementary Materials).

Target looks were reliably different from chance in clusters starting 500 ms after target onset in the control condition (z-sum = 93.02, $p < 0.001$) and 800 ms after target onset in the cohort condition (z-sum = 75.36, p < 0.001). These clusters started earlier than in the WebGazer data from the same participants, in which the corresponding clusters began 1100 ms after target

onset in both the control (z-sum = 38.65, $p < 0.001$) and cohort (z-sum = 35.28, $p < 0.001$) conditions.

Figure 4.8 shows looks to the competitor image in the cohort and control conditions. In the video data, competitor looks in the control condition decreased during target word articulation, whereas looks in the cohort condition did not decrease until target word offset. In contrast, in the WebGazer data from the same participants, competitor looks decreased only after target word offset in both conditions (similar to the pattern observed in the full WebGazer dataset), and competitor looks were more similar in the two conditions. In the video data, the analysis identified a reliable difference in competitor looks between conditions in a cluster 700–1099 ms after target onset (z-sum = 13.26, $p = 0.001$), thereby showing evidence of a phonemic cohort effect. A cluster analysis of the corresponding WebGazer data did not identify any clusters.

Figure 4.7: *Mean looks to the target and competitor images by condition in the Experiment 1A annotated webcam video data and in the WebGazer data from the same participants.* Ribbons indicate standard error. Vertical lines indicate average target word duration. Shading indicates when looks to the target image differed from chance.

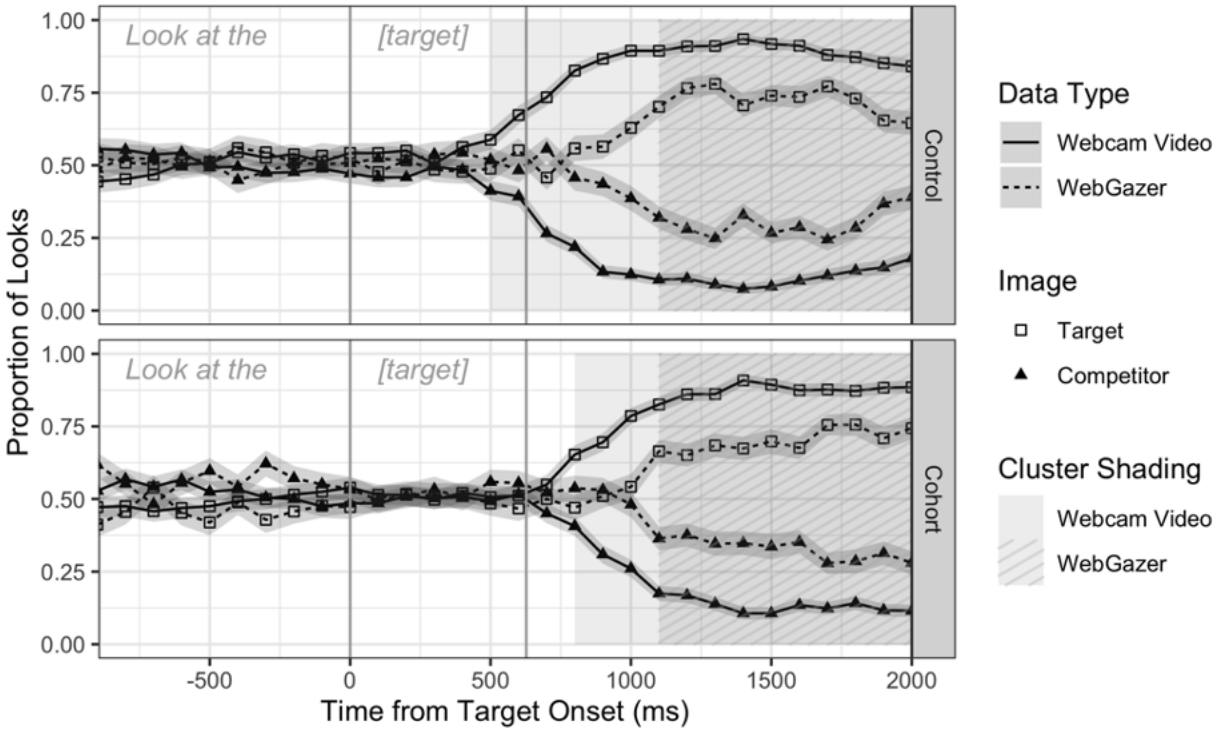
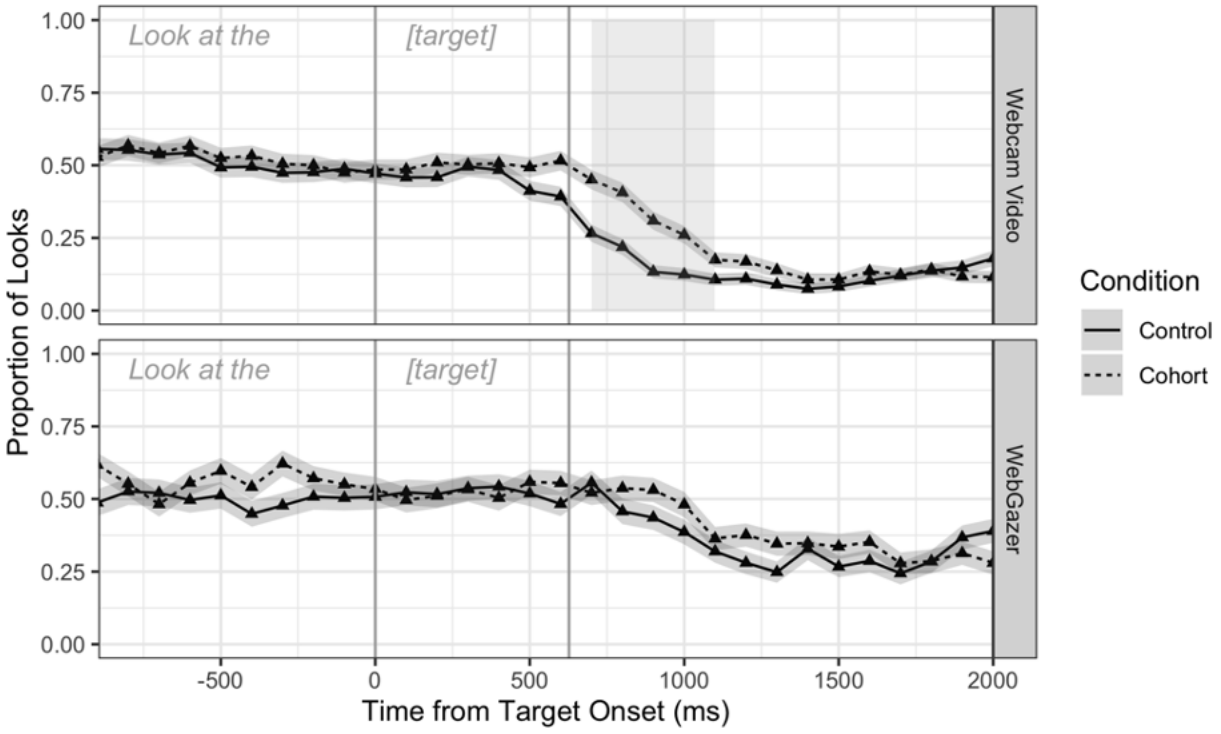


Figure 4.8: *Mean looks to the competitor image by condition in the Experiment 1A annotated webcam video data and in the WebGazer data from the same participants.* Ribbons indicate standard error. Vertical lines indicate average target word duration. Shading indicates when looks between conditions reliably differed.



4.2.4.2. Experiment 1B

Figure 4.9 plots looks to the target and competitor images in the cohort and control conditions, as identified by webcam video annotation and WebGazer for the same 13 participants (plots including distractor images are available in the Supplementary Materials). Target looks rose earlier and reached higher proportions in the video data than the WebGazer data.

In the video data for the control condition, looks to the side of the screen containing the target were reliably different from chance in clusters starting 600 ms after target onset along the horizontal axis (z-sum = 60.27, $p < 0.001$) and 700 ms after target onset along the vertical axis

(z-sum = 63.13, $p < 0.001$). In the cohort condition, clusters emerged 800 ms after target onset for the horizontal-side distinction (z-sum = 65.33, $p < 0.001$) and 600 ms after target onset for the vertical-side distinction (z-sum = 75.98, $p < 0.001$).

In the WebGazer data from the same participants, the detection of target looks appeared considerably later. In the control condition, target-side looks were reliably different from chance in clusters emerging 1200 ms after target onset along both the horizontal (z-sum = 34.73, $p < 0.001$) and vertical (z-sum = 15.96, $p = 0.001$) axes.[7] In the cohort condition, clusters emerged 1000 ms after target onset for the horizontal-side distinction (z-sum = 36.28, $p < 0.001$) and 1400 ms after target onset for the vertical-side distinction (z-sum = 15.80, $p = 0.001$).

Figure 4.10 shows participants' looks to the target and distractor images in the control condition 700–2000 ms after target onset (when participants were likely fixating on the target quadrant, according to the video annotation) for the webcam video data. The proportion of looks to the target was higher than during detected target fixations in the full WebGazer sample (Figure 4.5), and there were fewer distractor looks. Similar to the full WebGazer sample, there was a slight preference for vertical distractors over the other non-target images (this pattern was confirmed in an exploratory multinomial analysis; see Supplementary Materials) — however, the relative differences were smaller in the webcam video data. A figure showing target and distractor looks by target location is available in the Supplementary Materials.

Figure 4.11 shows looks to the competitor image in the cohort and control conditions. In the video data, looks to the competitor image in the cohort condition increased during target articulation, while looks in the control condition decreased. In the WebGazer data from the same

[7] The analysis of vertical-side looks identified two clusters: one 1200–1699 ms after target onset (z-sum = 15.96, $p = 0.001$) and one 1800–1999 ms after target onset (z-sum = 6.65, $p = 0.049$). The effect in the 1700 ms bin had a z-score of 1.98, so it did not meet the threshold to be included in a cluster.

participants, there was no obvious difference between conditions. In the video data, the analysis identified a reliable difference in competitor image looks between conditions in a cluster 600–999 ms after target onset (z-sum = 11.19, $p < 0.01$), thus finding evidence of a phonemic cohort effect. A cluster analysis of the corresponding WebGazer data did not identify any clusters.

Figure 4.9: *Mean looks to the target and competitor images by condition in the Experiment 1B annotated webcam video data and in the WebGazer data from the same participants.* Ribbons indicate standard error. Vertical lines indicate average target word duration. Shading indicates the temporal overlap of the clusters when target side looks differed from chance in both the horizontal and vertical directions.

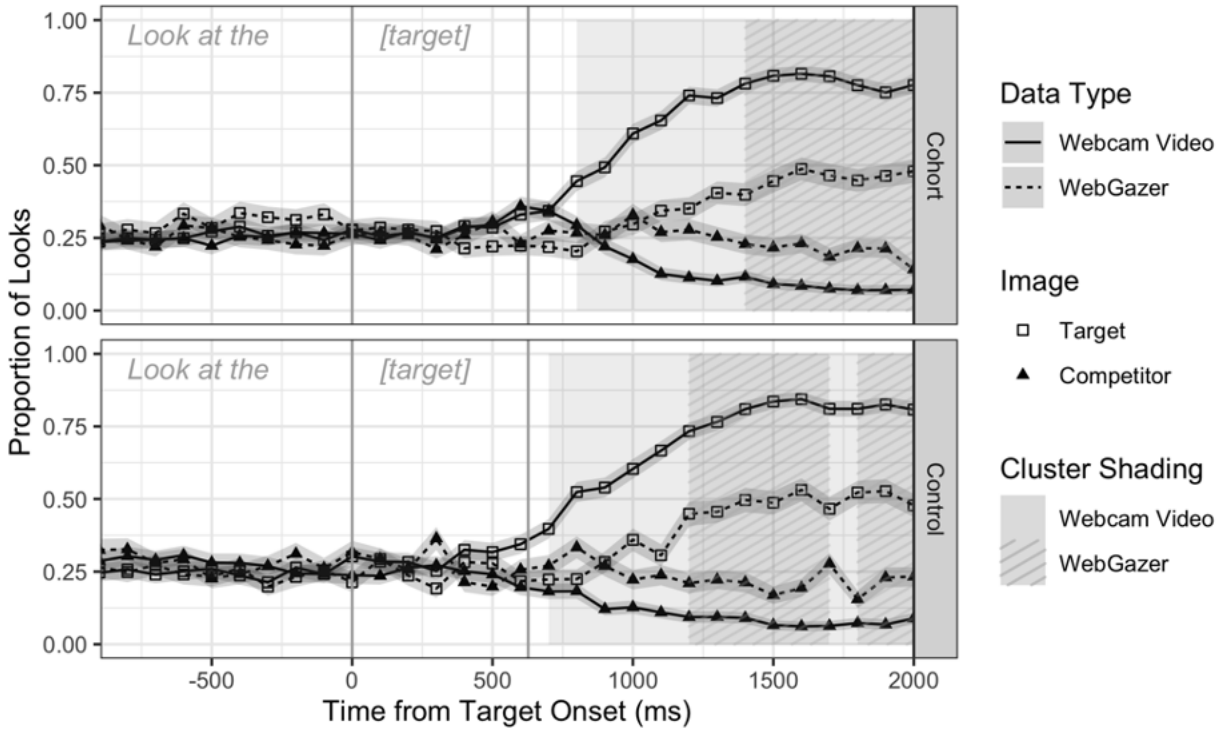


Figure 4.10: *Boxplot of participant fixation proportions to the target and non-target images in the Experiment 1B control trials from 700–2000 ms after target onset for the annotated webcam video data.* Mean fixation proportions for each image are labeled and identified by black diamonds. The gray points represent participant means.

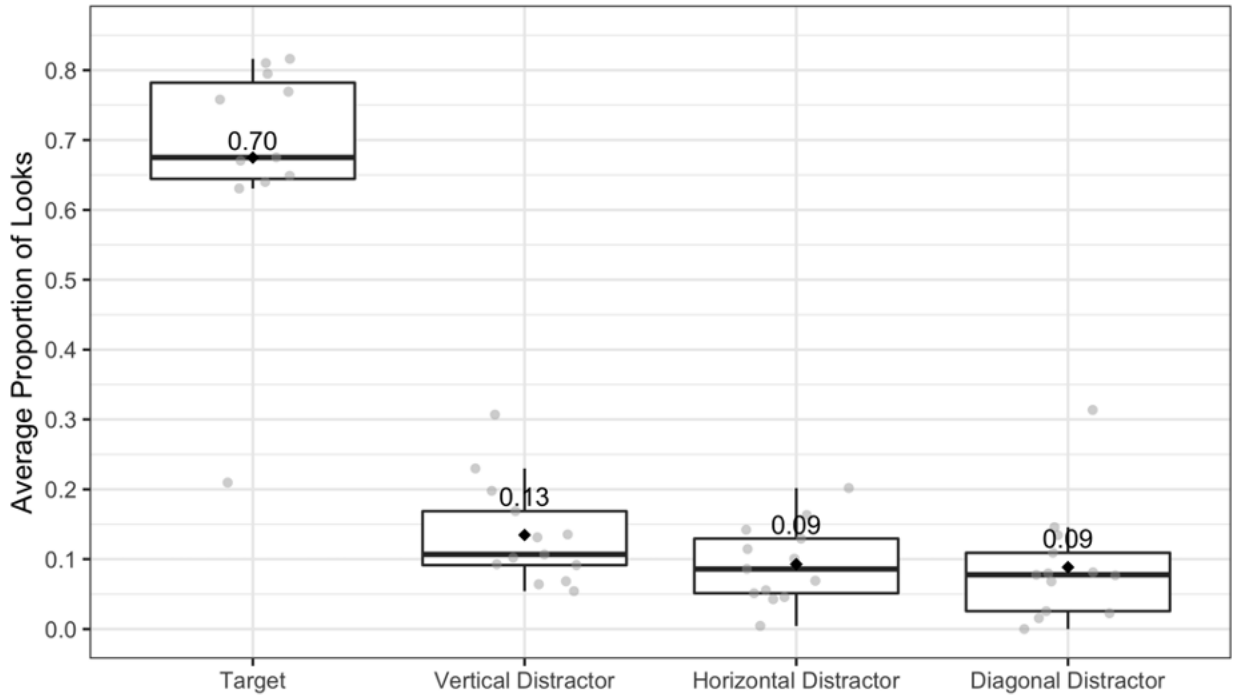
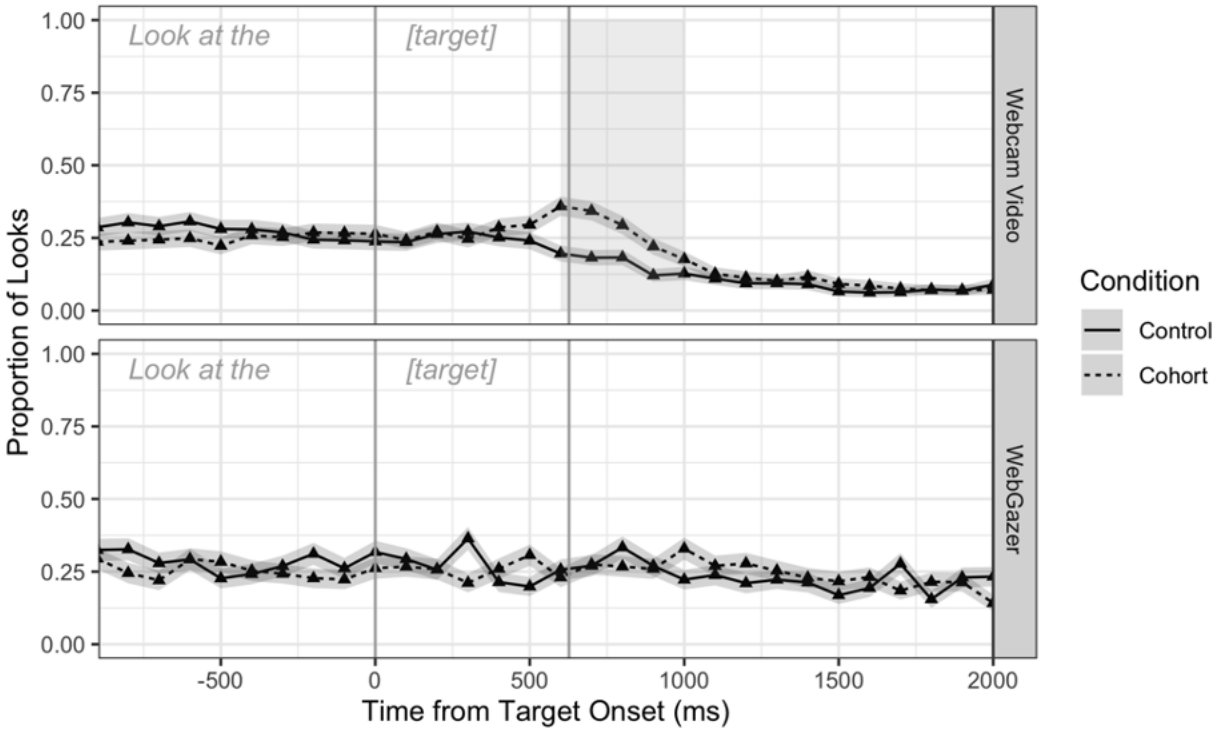


Figure 4.11: *Mean looks to the competitor image by condition in the Experiment 1B annotated webcam video data and in the WebGazer data from the same participants.* Ribbons indicate standard error. Vertical lines indicate average target word duration. Shading indicates when looks between conditions reliably differed.



4.2.3. Experiment 1 summary

Experiment 1 used a standard visual-world task to assess the relative performance of two webcam-based eye-tracking methods with five to six-year-old children: automatic WebGazer gaze coding and hand annotation of gaze direction from recorded webcam videos. Both methods detected increased looks to named (target) images in both two- and four-image displays. However, the rise in target fixations was lower and later in the WebGazer data compared to in-lab experiments with children of the same age or younger (e.g., Sekerina & Brooks, 2007; Simmons, 2017). The annotated video data, on the other hand, looked more like data collected in

in-lab experiments: The onset of target looks was faster, and the proportion of target looks was considerably higher than in simultaneously-collected WebGazer data. Interestingly, for both methods, unrelated images vertically-adjacent to the target received more looks than distractor images in the other locations of the display; this pattern was especially notable in the WebGazer data.

The differences between the two methods were particularly pronounced in the analysis of the phonemic cohort effect. In the video data, the cohort effect emerged in both the four- and two-image displays in clusters beginning 600–700 ms after target onset and was detectable in a sample of just 13 children. This effect is later than observed in previous lab-based studies, in which cohort effects began 200–400 ms after target onset (e.g., Allopenna et al., 1998; Huettig & McQueen, 2007; Sekerina & Brooks, 2007). While this difference could reflect our small sample size or a difference in our analysis method, it is consistent with other research using webcam video annotation (i.e., the web-based replication of Allopenna et al., 1998 by Ovans, 2022). In the WebGazer data, the effect was detectable only in the two-image display with a larger sample ($N = 32$), and this effect window emerged later (900 ms after target onset). These results suggest that while WebGazer can detect robust fixation patterns like target looks, webcam video annotation is better suited to detecting more fine-grained effects.

In Experiment 1 we tracked looks in a binary fashion, monitoring whether or not a look fell inside a particular region. While this measure reflects how visual-world studies are generally conducted, we cannot tell from these results how close WebGazer's gaze estimates are to the true locations of visual stimuli. Experiment 2 explores WebGazer's accuracy more directly. This additionally allows us to address one limitation of Experiment 1: Because our image canvases did not cover the full halves (Experiment 1A) or quadrants (Experiment 1B) of the screen, gazes

that were estimated to fall near a canvas, but not within it, may have been coded as looks in our video data but not in the WebGazer data.

4.3. Experiment 2: Fixation task

Experiment 2 used a visual-fixation task to investigate the spatial and temporal resolution of WebGazer's gaze estimation with four to twelve year-old children. This task was adapted from Slim and Hartsuiker (2022) ("S&H2022"). The experiment had four goals: (i) to assess the feasibility of conducting web-based eye-tracking tasks with children without an experimenter present; (ii) to assess how closely WebGazer estimates correspond to stimulus locations; (iii) to assess whether there are age-related differences in WebGazer performance between four and twelve years; and (iv) to assess whether the accuracy of quadrant-based analyses with WebGazer is improved by using larger canvases.

4.3.1. Methods

4.3.1.1. Participants

The study included 45 participants between four and twelve years of age (Table 4.1). Participants spoke American or British English natively. Participants were not required to be monolingual, as the experiment was non-linguistic. Three participants in Experiment 2 previously took part in Experiment 1 during a different experiment session. Informed written consent was received from the parent or guardian for their child's participation; child participants additionally provided written assent. Participants were compensated with a $5.00 gift card.

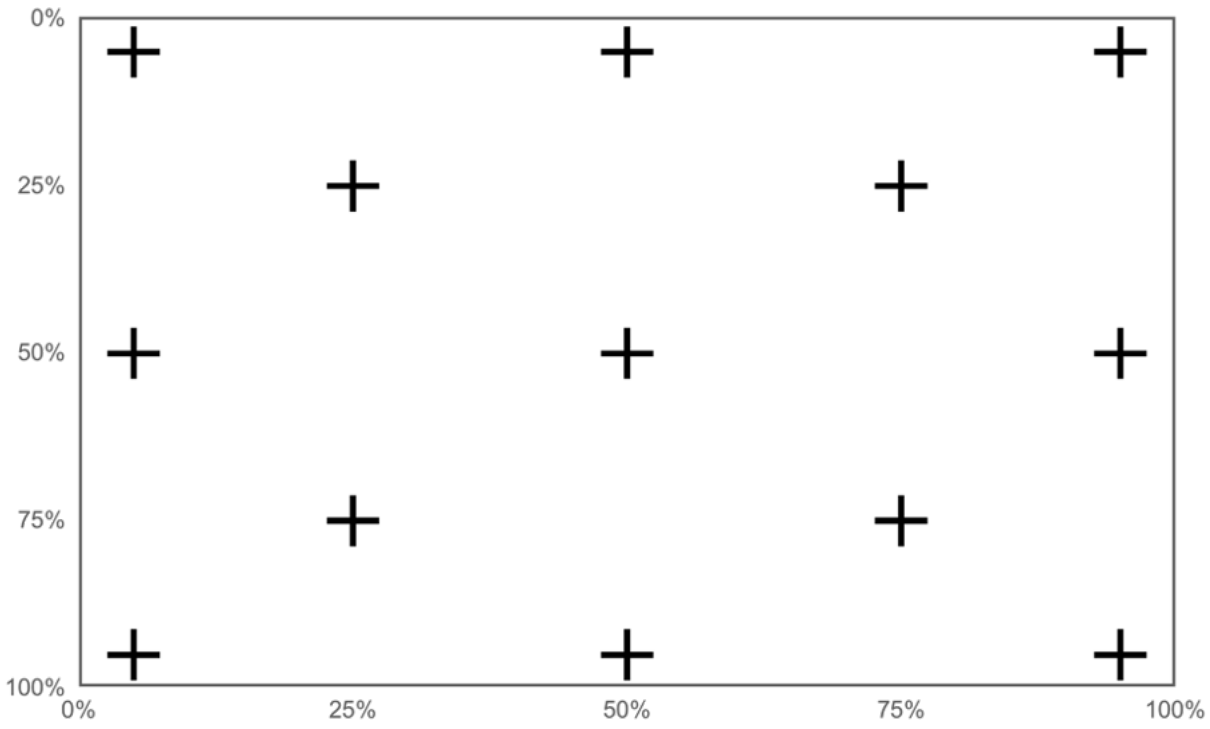
Table 4.1: *Experiment 2 participant ages.*

Age group	Count	Mean age (months)	Range (years;months)
4–5 years	12 (7 F, 5 M)	62.9 (SD = 5.2)	4;6–5;9
6–7 years	12 (5 F, 7 M)	81.8 (SD = 6.6)	6;0–7;7
8–9 years	11 (4 F, 6 M, 1 NB)	104.4 (SD = 4.5)	8;3–9;6
10–12 years	10 (7 F, 3 M)	132.5 (SD = 9.6)	10;0–12;8

4.3.1.2. Materials

The experiment was built in PCIbex (Zehr & Schwarz, 2018) using WebGazer v2 and was completed in the participant's web-browser. The stimuli were modeled on those from S&H2022. Participants looked to fixation crosses that appeared in 13 possible screen positions (Figure 4.12). Each fixation cross appeared in each location six times, resulting in 78 total trials. Trial order was randomized for each participant. To accommodate variability in computer screen-sizes, the experiment was completed in fullscreen, and stimulus size and location were defined by browser window size. To make the task more fun for child participants, the 78 trials were divided into six blocks: in each block, the fixation cross appeared in a different color and was accompanied by a different audio sound effect.

Figure 4.12: *The 13 possible target stimulus locations in Experiment 2.* The panel represents the full experiment screen (the axis labels indicate percentage of screen-size).



4.3.1.3. Procedure

The experiment was completed by participants from their own computers, unsupervised by researchers. An experiment access link was sent to the parent's email. Participants were asked to complete the experiment on a computer or laptop using either Google Chrome or Mozilla Firefox. An adult was asked to help the child get set up and to remain in the room as they completed the task.

The experiment started with an introductory sequence that walked participants through an audio check, the WebGazer calibration, and the experiment instructions. Following the audio check, the sequence included both written and auditory instructions so that it would be accessible

to both child participants and adult supervisors. As in Experiment 1, we did not specify a minimum calibration threshold. Participants were instructed to look at the plus signs that appeared on the screen; they were instructed to look at them as fast as they could and to stare at them until they disappeared.

To start each block of the task, the participant pressed the spacebar, which initiated a calibration check (resulting in seven total calibration scores per participant). The trial structure was the same as in S&H2022. Each trial began with a small black fixation cross (+) appearing in the center of the screen for 500 ms (font size defined as 5% of the screen height). This cross then disappeared and the colored target fixation cross appeared on screen for 1500 ms (size defined as 10% of the screen height). The trial then ended, and the next trial began automatically. The experiment took approximately 10–15 minutes to complete.

4.3.2. Data processing

WebGazer tracked participants' eye-movements from target stimulus onset to trial offset. We recorded looks to canvases covering each quadrant of the screen as a binary variable. These canvases together covered the entire screen (each 50% of browser window height and width). We also recorded coordinate estimates of gaze location (in pixels). If either the x- or y-coordinate estimate was missing in the recorded WebGazer data, the sample was omitted from the data prior to processing (0.53% of samples).

Data processing followed the procedure outlined by S&H2022 using the scripts made available in their OSF repository (https://osf.io/yfxmw/). We aggregated the data into 100 ms bins, calculating for each bin the mean x- and y-coordinate estimates and mean looks to each quadrant canvas (quadrant looks were later binarized for analysis). We restricted the dataset to

bins ranging from 0–1500 ms after trial onset, resulting in exclusion of 170 out of 3484 recorded bins (4.88%). To account for participants' different screen-sizes, we converted the pixel coordinate estimates to a distance metric based on screen-size proportion, such that the pixel in the center of the screen had coordinates (0.5, 0.5) and the pixel in the bottom right corner had coordinates (1,1). For each bin, we calculated the Euclidean distance between the estimated gaze location and the center of the target fixation cross (in proportion of screen-size) using the formula below.

$$\sqrt{\left(x_{target} - x_{gaze\ estimation}\right)^2 + \left(y_{target} - y_{gaze\ estimation}\right)^2}$$

In some of our analyses, we compare the Experiment 2 data to S&H2022's adult data accessed from their OSF repository.

4.3.3. Results

26 trials across 10 participants were omitted from the analysis because no WebGazer data were saved for them on our server.

4.3.3.1. Calibration scores

Participant calibration scores in the initial calibration sequence ranged from 6–80%, with an average score of 52% (SD = 16). The mean participant scores for the six calibration checks ranged from 4–67%, with an average of 39% (SD = 13). Table 4.2 summarizes participant mean calibration scores (calculated using all seven scores for each participant) by age group.

As in S&H2022, participant mean calibration scores were significantly correlated with webcam sampling rates (measured in frames per second) ($\rho = 0.33$, $p = 0.03$), suggesting that WebGazer's estimates are more precise when there are more recorded samples. In addition,

mean calibration score was significantly correlated with participant age in months (to one decimal place) ($\rho = 0.52$, $p < 0.001$), indicating that older participants tended to have higher calibration scores. This trend still holds when accounting for sampling rate (see Supplementary Materials for more details and plots).

Table 4.2: *Experiment 2 mean participant calibration scores by age group.*

Age group	Mean score (%)	Range (%)
4–5 years	33 (SD=13)	11–50
6–7 years	40 (SD=12)	4–52
8–9 years	43 (SD=9)	28–52
10–12 years	51 (SD=10)	35–66

4.3.3.2. Euclidean distance from the target over time

To assess how closely WebGazer estimates match stimulus location, we plotted the mean Euclidean distance (in percentage of screen-size) between the target stimulus and estimated gaze location from stimulus onset to trial offset (Figure 4.13). The plot includes data for the Experiment 2 child participants as well as S&H2022's adult participants. In both populations, distance from the target began to decrease 200 ms after stimulus onset and plateaued around 500 ms after onset. While this timing is similar for the two populations, the Euclidean offset was larger and more variable for children, settling at an offset of approximately 38% of screen distance from the target.

To better understand the factors influencing this offset, we plotted mean distance over time broken down by calibration score (Figure 4.14) and child age group (Figure 4.15).[8] Figure 4.14 illustrates the relationship between mean calibration score and WebGazer's spatiotemporal accuracy: mean Euclidean offset was smaller for participants in higher calibration bins. Figure 4.15 suggests that there was also a relationship between Euclidean distance and age: Offsets plateaued at the shortest distance for the 10–12 year-old participants, followed by the 8–9 year-old and 6–7 year-old participants, with the longest distance offsets for the 4–5 year-old participants. However, as discussed above, mean calibration score and age were correlated; therefore, it is not obvious from Figure 4.15 the extent to which age contributed to Euclidean offset independently from calibration score. We address this below.

[8] Given variations in WebGazer sampling, there were fewer samples towards the end of the trial (see also S&H2022). We thus ended the plots at 1200 ms, the latest time point for which we had enough samples to calculate standard errors in all bins for both plots.

Figure 4.13: *Mean Euclidean distance (in percentage of screen-size) from the target stimulus over the course of the trial.* Error bars indicate standard deviation. Ribbons indicate standard error.

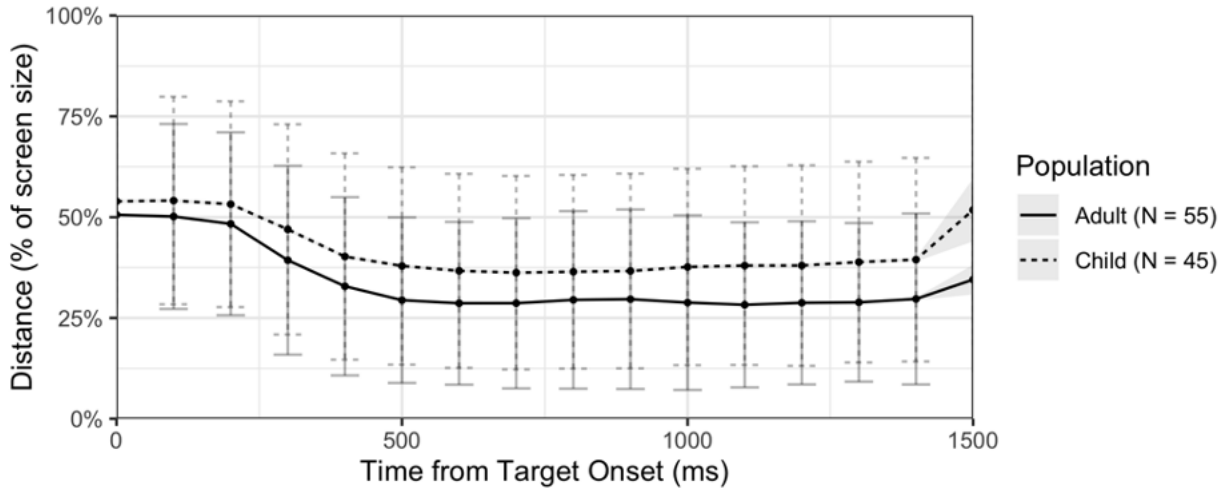


Figure 4.14: *Mean Euclidean distance (in percentage of screen-size) from the target stimulus over the course of the trial, broken down by participant calibration score.* Ribbons indicate standard error.

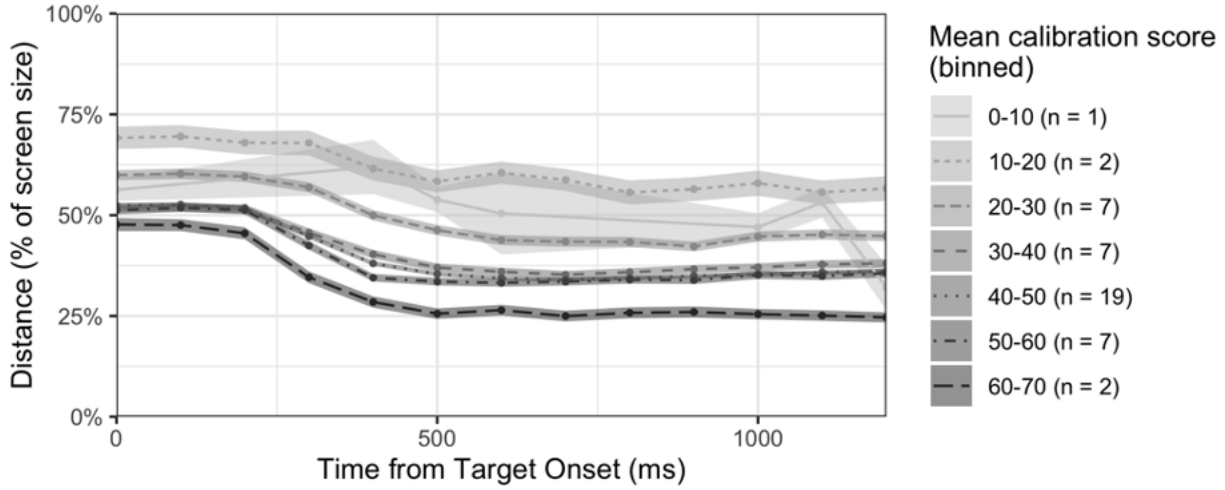
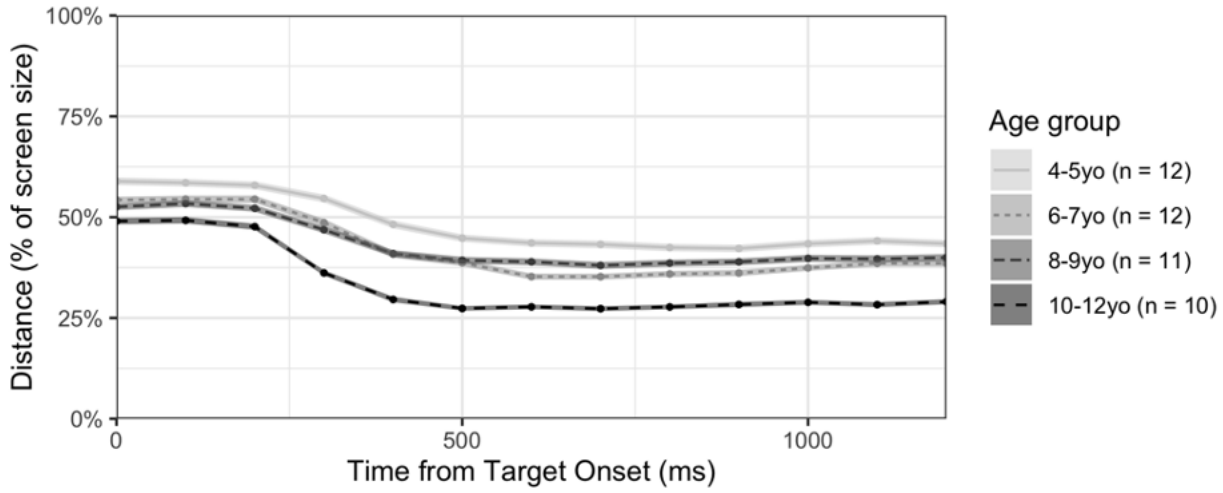


Figure 4.15: *Mean Euclidean distance (in percentage of screen-size) from the target stimulus over the course of the trial, broken down by participant age bin.* Ribbons indicate standard error.



4.3.3.3. Looks in the fixation window

As in S&H2022, we analyzed a fixation time window 500–1500 ms after target onset to assess WebGazer's spatial resolution when gaze had settled on the target location. Figure 4.16 plots the density of looks on the screen during this time window for all 13 target locations. Density plots of the quadrant fixations for the youngest and oldest age groups in our sample are available in the Supplementary Materials.

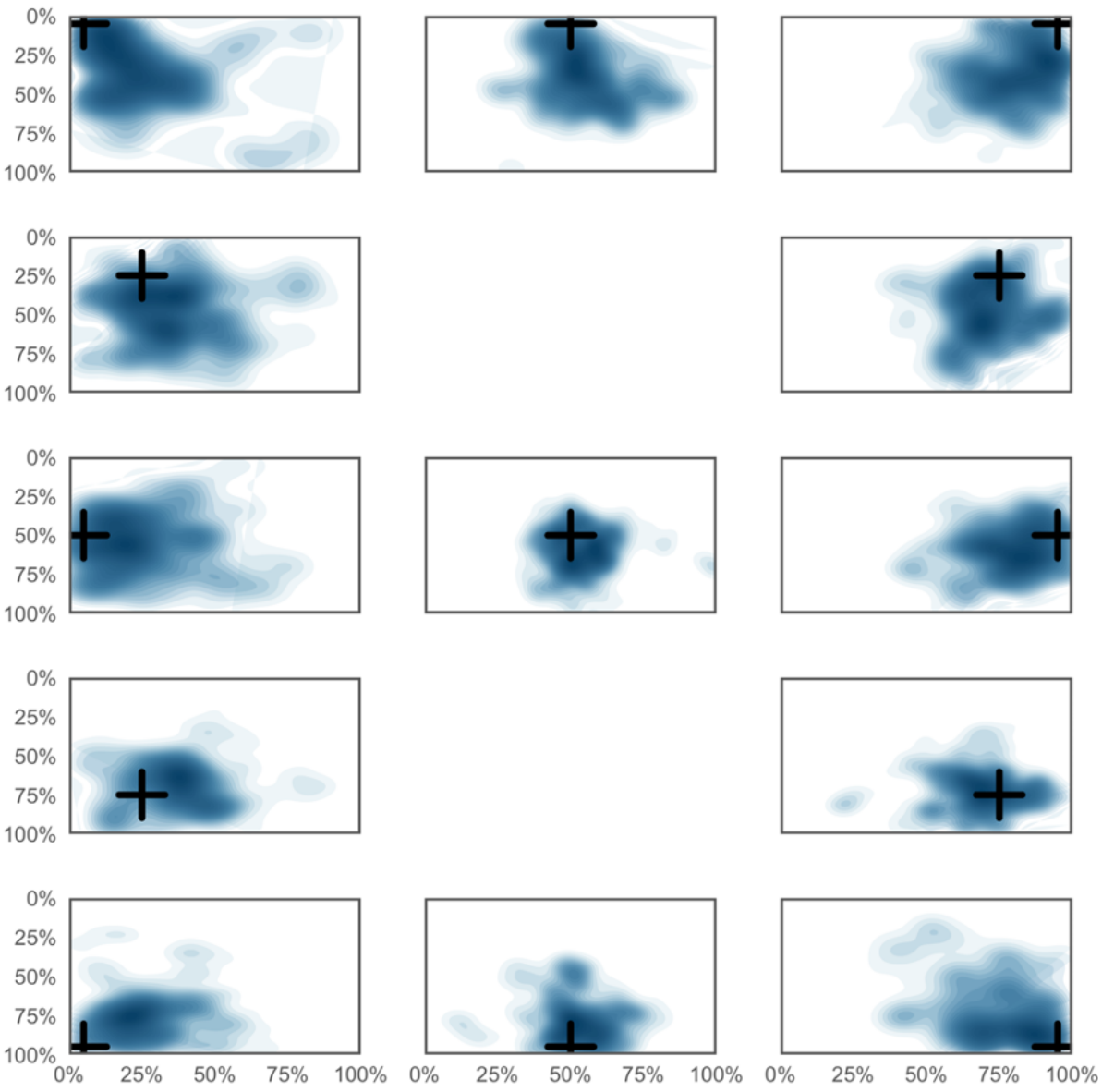
For each location, estimated looks tended to fall around the stimulus, though the range in which the looks fell was large. In the plots for the targets appearing in the center of each quadrant (the second and fourth row of Figure 4.16), estimated looks often extended into quadrants other than the one containing the target stimulus, with particular overlap in the vertical direction. In fact, within the fixation window, participants' mean vertical offsets between their estimated gaze location and the true stimulus location (M = 0.27, SD = 0.08) were greater than their mean horizontal offsets (M = 0.21, SD = 0.08) ($t(44) = 5.55$, $p < 0.0001$). WebGazer's

reduced vertical accuracy appears particularly pronounced in the upper quadrants of the screen (the second row of Figure 4.16).

To assess the relative contributions of calibration accuracy and participant age to Euclidean distance offset during target fixations, we calculated the mean distance from the target during the 500–1500 ms fixation window for each participant and computed a linear regression with fixed effects of mean calibration score and age in months (to one decimal place). Both the effects of mean calibration score (β = -0.004, $t(42)$ = -3.93, $p < 0.001$) and age (β = -0.001, $t(42)$ = -2.30, $p = 0.03$) were reliable.[9] Model comparison using ANOVA revealed significant differences between models with both predictors and models with only calibration score ($F(1,42)$ = 5.27, $p = 0.03$) and only age ($F(1,42)$ = 15.4, $p < 0.001$). These results suggest that there was an effect of age on Euclidean offset that was distinct from the effect of calibration score.

[9] Given the correlation between the two model predictors, multicollinearity in the model was assessed by calculating the Variance Inflation Factor (VIF); VIF for both predictors was 1.38.

Figure 4.16: *Density plots indicating estimated looks on the screen 500–1500 ms after target onset for each possible target location.* Each panel represents the full experiment screen (the axis labels indicate percentage of screen-size), and the black crosses indicate the center of the target locations.



4.3.3.4. Quadrant looks over time

In addition to investigating Euclidean distance over time, we also analyzed quadrant looks over time, allowing us to assess WebGazer's accuracy discriminating quadrant looks when using larger canvases than in Experiment 1B. We restricted the data to the trials in which the fixation cross appeared in the center of each screen quadrant. We binarized quadrant looks using the same procedure as in Experiment 1B. For comparison, we also binarized S&H2022's adult data in the same fashion.

Figure 4.17 plots looks to the target quadrant over time compared to the other quadrant locations (horizontally, vertically, or diagonally across from the target) for both populations. A plot showing quadrant looks by target location for the child participants is available in the Supplementary Materials. The pattern of target quadrant looks in the child data resembles that observed in Experiment 1B: Looks to all quadrants began at chance (25%), and then looks to the target increased and plateaued around 50%.

To assess target quadrant looks, we performed the same target side analyses as we conducted for Experiment 1B. We analyzed looks from 0–1400 ms after target onset given the reduced number of samples at the end of the trial. The analyses followed the same procedure as the Experiment 1B analyses, except quadrants were used instead of images, and item was defined as target location (top left, top right, bottom left, bottom right). In the Experiment 2 child data, looks to the side of the screen containing the target were reliably different from chance in clusters starting 300 ms after target onset along both the horizontal (z-sum = 92.71, $p < 0.001$) and vertical (z-sum = 56.30, $p < 0.001$) axes (in the exploratory multinomial analysis, target quadrant looks similarly differed from looks to all other quadrants in a cluster starting 300 ms after onset; see Supplementary Materials). In the S&H2022 adult data, clusters started at target

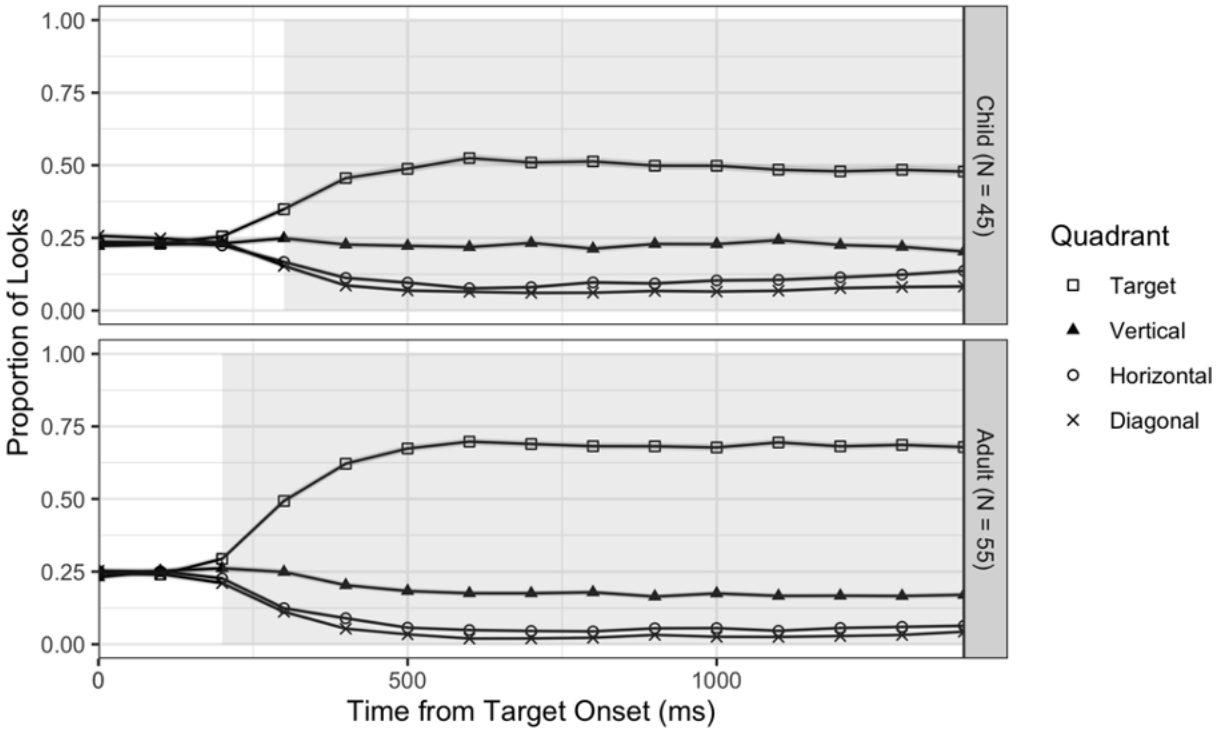
onset for the horizontal-side distinction (z-sum = 136.59, $p < 0.001$)[10] and 200 ms after target onset for the vertical-side distinction (z-sum = 90.31, $p < 0.001$), suggesting that WebGazer detected target quadrant fixations starting 200 ms after target onset (in the exploratory multinomial analysis, looks to the target differed in a cluster starting 300 ms after onset).[11]

In both the adult and child data, looks to the quadrant vertically adjacent to the target remained elevated compared to the other non-target quadrants during target fixations (Figure 4.17). This pattern was confirmed in an exploratory multinomial analysis (see Supplementary Materials).

[10] The early cluster onset identified in the horizontal target-side looks analysis appears to be driven by a slight preference for target-side looks in the 0 ms and 100 ms time bins (in both bins, 54% of recorded looks fell on the target side). The beta coefficients for these time bins (0.17 and 0.12, respectively) suggest that the effect was small; for reference, the beta coefficient 500 ms after target onset (once target looks plateau in Figures 4.13 & 4.15) was > 2.

[11] This timing differs from that identified in S&H2022's analysis, which identified clusters in the 0–200 ms and 400–1400 ms bins (Slim & Hartsuiker, 2022). Their analysis collapsed all non-target quadrants into a single other quadrant variable that received a 1 if there was a look to any quadrant other than the target quadrant; they then analyzed whether looks (0,1) differed based on focus (target quadrant, other quadrant). We did not perform our analyses this way due to concerns about dependencies in the data structure. Repeating our cluster analysis using this structure, we obtained the same two clusters identified by S&H2022 (in the bins 0–200 ms and 400–1400 ms after target onset). Using S&H2022's analysis structure on the Experiment 2 child data yields clusters in the 0–300 ms and 500–1400 ms bins.

Figure 4.17: *Quadrant looks over time for the Experiment 2 child participants and Slim and Hartsuiker's (2022) adult participants.* Ribbons indicate standard error. Shading indicates the temporal overlap of the clusters when target side looks differed from chance in both the horizontal and vertical directions.



4.3.4. Experiment 2 summary

The Experiment 2 results demonstrate that it is possible to conduct unsupervised web-based eye-tracking tasks with school-aged children. There was a sharp increase in looks toward the target shortly after it appeared, indicating that participants were able to perform the task without an experimenter to guide them. Furthermore, the data suggest that parents, acting on their own, were just as effective in setting up the experiment as parents guided by researchers; there was no significant difference between the mean calibration scores of participants in

Experiment 1 and those of the same age range (five–six years) in Experiment 2 ($t(7) = -0.74$, $= 0.49$).

Experiment 2 assessed how closely WebGazer estimates track with stimulus location. The Euclidean offset between estimated gaze and target stimulus location was approximately 38% of screen-size. This offset is greater than observed in S&H2022's adult data (30% of screen-size) and much larger than reported for in-lab eye-trackers (see *§4.4. General Discussion*). In addition, we found age-related differences in performance: Calibration scores tended to be higher for older participants, and there was a relationship between participant age and Euclidean offset above and beyond the effect of calibration.

Analyzing the Experiment 2 data using the quadrant-based approach common for visual-world studies showed a similar overall pattern to Experiment 1B, suggesting that increasing the size of the tracked quadrant canvases did not substantially improve WebGazer data quality. Target quadrant looks increased faster in Experiment 2 than Experiment 1, likely due to the larger quadrant sizes and the fact that attention was directed to the stimulus by a single visual cue (the target was the only item on screen), whereas in Experiment 1 participants needed to process linguistic input to determine which of multiple visual stimuli to fixate upon. Our results furthermore support the finding from Experiment 1 that WebGazer is less accurate at detecting vertical distinctions: During target fixations, offsets between estimated gaze locations and the true stimulus location were greater in the vertical direction than the horizontal direction, and in the quadrant-based analysis, there were elevated looks to the quadrant vertically adjacent to the target. This inaccuracy appears to be particularly pronounced in the top–down direction, with greater vertical offsets for targets appearing on the top half of the screen (Figure 4.16).

4.4. General discussion

The present study investigated the suitability of two webcam eye-tracking methods for child language research: automatic WebGazer gaze estimation and frame-by-frame annotation of gaze direction from webcam videos. Experiment 1 compared these two methods with five and six year-olds in a visual-world task replicating the phonemic cohort effect. The experiment used two display types: a two-image display with one image on each side of the screen (Experiment 1A) and a four-image display with one image in each quadrant (Experiment 1B). Experiment 2 investigated WebGazer's gaze estimation accuracy in an unsupervised visual-fixation task with four to twelve year-old children. Our results suggest that while it is possible to conduct webcam eye-tracking studies with children (supervised and unsupervised), the two eye-tracking methods differ in their spatiotemporal resolution and thus are not equally suitable for detecting all types of eye-movement patterns. In this Discussion, we discuss the spatiotemporal accuracy of the two methods, their ability to detect fine-grained linguistic effects, and recommendations for researchers conducting web-based eye-tracking experiments with children.

4.4.1. Spatiotemporal accuracy of the eye-tracking methods

4.4.1.1. Spatial resolution

Both webcam eye-tracking methods were sufficiently accurate to detect the preference to look at a target that either is explicitly mentioned (Experiment 1) or suddenly appears on the screen (Experiment 2). This is true both when target and foil occupy different halves of the screen (Experiment 1A) and when the target occupies one quadrant (Experiment 1B, Experiment 2). Nevertheless, webcam video annotation had a higher signal-to-noise ratio, as evidenced by a

higher proportion of target looks than in simultaneously-collected WebGazer data (89% vs. 72% in Experiment 1A; 80% vs. 47% in Experiment 1B). In fact, the target looks in the video data parallel those from prior in-lab experiments using commercial eye-trackers with children of this age (Sekerina & Brooks, 2007; Simmons, 2017).

Experiment 2 confirmed WebGazer's reduced spatial accuracy compared to in-lab eye-tracking using a more fine-grained distance metric. Target fixations as detected by WebGazer were approximately 38% of the screen distance from the true stimulus location, compared to an offset of 1–2% (0.4–0.9º of the visual angle) reported for Tobii TX300 eye-trackers in standard laboratory conditions (Tobii, 2010; see Dalrymple et al., 2018 for data from 8–11 year-old children). This offset could result from WebGazer inaccuracy or because participants are looking somewhere else. We believe that the latter explanation does not play a substantial role, as: (i) piloting the task over Zoom showed children directing their eyes towards target stimuli; (ii) children similarly directed their eyes towards visual cues in the Experiment 1 WebGazer calibration sequence; and (iii) if inattention were the primary driver of this gap, we would expect to see a more random distribution of looks in the Figure 4.16 density plots. Furthermore, this larger offset is consistent with prior WebGazer studies with adults; for example, Semmelmann and Weigelt (2018) and Slim and Hartsuiker (2022) reported offsets of 18% and 30% of screen-size (respectively) in online fixation tasks.

In particular, WebGazer appears to have difficulty discriminating looks along the vertical axis. This was evidenced by elevated looks to the image or quadrant vertically adjacent to the target and by greater vertical than horizontal offsets between gaze and target location. The results of Experiment 2 suggest that this difficulty may be greater for stimuli appearing on the top half of the screen. We observed a similar (but less pronounced) pattern in the video data in

Experiment 1. Poor vertical resolution could reflect three constraints. First, most computer screens are rectangles with a landscape orientation, thus vertical distances between stimuli are generally smaller than horizontal ones. Second, webcams are typically placed above the screen but centered on the left–right axis. Consequently, a left look will be in the opposite direction relative to the webcam from a right look. In contrast, looks to both the upper and the lower half of the screen will be downward relative to the webcam. Finally, while it is easy to encourage participants to center themselves relative to their screen on the left–right axis (by sliding their computer or chair), vertical position is variable and more difficult to control. Most adults sit with their eyes above the screen, and thus the WebGazer algorithm was presumably trained on data of this kind. Children, who are shorter but live in a world of artifacts scaled to adults, typically sit with their heads nearer to the level of the screen. This may explain why the vertical spread is greater for children in the WebGazer data.

In our study, we identified two factors that influence WebGazer's performance: calibration score and participant age. Higher calibration scores are associated with data patterns suggesting better gaze tracking. In Experiment 2, the distance between estimated looks and the true stimulus location was reduced for participants with higher mean calibration scores (see also Slim & Hartsuiker, 2022). In both Experiment 1 and the adult pilot experiment, the size of the cohort effect was larger in trials with higher scores on the preceding calibration check (see Supplementary Materials). These results highlight the potential utility of calibration thresholds as a means to improve data quality, though the threshold of 50% often used in adult WebGazer experiments (e.g., Slim & Hartsuiker, 2022, Experiment 2; Vos et al., 2022) may be too high a bar for younger child participants (see Table 4.2).

Participant age also seems to influence WebGazer accuracy. WebGazer's spatial resolution appears higher for adult participants compared to child participants: In Experiment 2, the Euclidean distance offset between estimated gaze location and the true target location was smaller for adults, and in the quadrant analysis, the adult data yielded a higher proportion of target quadrant looks. Moreover, the age of the child participants influenced both calibration score and Euclidean distance offset: Calibration scores were higher and distance offsets were smaller for older children.

Age-related differences could reflect factors specific to WebGazer. For example, older children may be in a more optimal position for WebGazer, because they are generally taller and thus may be positioned more like adults. In addition, older children tend to have larger faces than younger children, which could facilitate WebGazer's pupil detection and gaze estimation algorithms. Alternatively, age-related differences could reflect differences between participants that are independent of the technology used to estimate gaze. For example, older children may be less susceptible to distraction and more likely to sit still throughout the duration of the task.

4.4.1.2. Temporal resolution

In addition to observing differences in the eye-tracking methods' spatial resolutions, we also found that the timing of effects was slower than expected when WebGazer was used. In Experiment 1, in the absence of cohort competition, WebGazer detected reliable preferences for the target in clusters starting 800 ms after target word onset for the two-picture display and 1200 ms after target onset for the four-picture display. In contrast, in the annotated video data, this preference emerged in clusters starting 500 ms after target word onset in the two-picture display and 700 ms after target word onset in the four-picture display, similar to the timing in laboratory-

based studies (Sekerina & Brooks, 2007; Simmons, 2017). We also found delays in the timing of the phonemic cohort effects (discussed below).

The apparent lag is not limited to studies with linguistic stimuli: We observed comparable WebGazer fixation delays in Experiment 2. In in-lab settings, saccade latencies in response to perceptual stimuli take approximately 200–250 ms for adults (e.g., Matin et al., 1993; Rayner, Slowiaczek et al., 1983; Saslow, 1967; Theeuwes et al., 1998; Walker et al., 2000; White et al., 1962) and ten to twelve year-old children (Yang et al., 2002). Yang et al. (2002) observed mean latencies of approximately 300–350 ms for children between the ages of four-and-a-half to twelve years. In contrast, in Experiment 2, looks settled on the target location approximately 500 ms after onset (see also Semmelmann & Weigelt, 2018; Slim & Hartsuiker, 2022, for evidence of fixation delays with WebGazer).

We can imagine two possible explanations for this lag, which are not mutually exclusive. First, WebGazer could detect the same eye-movements as other eye-tracking measures but do so later due to time-consuming steps in the execution of the algorithm. Second, the lag could be a side-effect of WebGazer's poorer signal-to-noise ratio: Effect sizes at the onset of an eye-movement pattern are typically smaller, making differences more difficult to detect. The data to date suggest that both factors play a role. On the one hand, streamlining WebGazer's algorithm to remove unnecessary computations improves its temporal resolution (Yang & Krajbich, 2021), suggesting processing limitations result in temporal delays. On the other hand, the variability that we observed in the WebGazer estimates well after stimulus onset (Figure 4.16) demonstrates that the spatial signal has substantial noise. Since even more streamlined versions of the WebGazer algorithm produce smaller effects than in-lab baselines (Vos et al., 2022), we expect that they would also fail to detect the earliest and weakest effects. Critically, we did not see comparable

delays in the simultaneously-collected video data (Experiment 1), demonstrating that these delays are due to properties of the WebGazer algorithm and its execution and not to the less controlled nature of web-based settings.

4.4.2. Using webcam eye-tracking to detect fine-grained linguistic effects

Our findings suggest that WebGazer is not well suited for studying small or fleeting effects in children, particularly in the typical quadrant-based visual-world display. This was most clearly demonstrated by our analyses of the phonemic cohort effect in Experiment 1. In the annotated webcam video data, we found significant cohort effects in both the two- and four-image displays, despite a sample of just 13 participants in each experiment. In contrast, even though our WebGazer sample contained more than twice as many participants ($N = 32$ per experiment), WebGazer only detected evidence of a cohort effect in the two-image display. Moreover, the cluster window containing the effect was later and shorter than that in the video data (extending from 900–1099 ms vs. 700–1099 ms). Prior studies with adults have similarly observed WebGazer effects emerging later than in-lab baselines (Degen et al., 2021; Slim & Hartsuiker, 2022) as well as effects that are smaller and/or noisier than in-lab counterparts (Degen et al., 2021; Vos et al., 2022; Slim & Hartsuiker, 2022).

We conducted a series of power simulations using the {mixedpower} package v0.1.0 (Kumle et al., 2021) to assess the relative effect sizes in our webcam video and WebGazer data (see Supplementary Materials). We analyzed the likelihood of competitor image looks in the time windows where the Experiment 1 cluster analyses identified a difference between the two conditions (700–1099 ms in Experiment 1A; 600–999ms in Experiment 1B).

In Experiment 1A (the two-image display), the effect of condition was larger in the webcam video data (β = -1.12, z = -3.52, $p < 0.001$) than in the WebGazer data (β = -0.29, z = -1.83, p = 0.07). In fact, the WebGazer effect was only 26% as large as the webcam video effect (as measured by the standardized beta coefficients). Given the sample sizes that we had, the observed power was 94% in the video data (N = 13) and 45% in the WebGazer data (N = 32). To achieve 94% power with the WebGazer data, the sample size would have to be increased to approximately 125 participants. To achieve power of at least 80% in the WebGazer data, the sample size would have to be increased to approximately 65 participants (power = 80%). In contrast, reaching 80% power in the video data requires only seven participants (power = 81%). In short, these simulations suggest that to achieve comparable power, a WebGazer study of this kind would require almost ten times as many participants as a study relying on webcam video annotation.

In Experiment B (the four-image display), the effect of condition was significant in the webcam video data (β = -1.08, z = -3.97, $p < 0.0001$), with an observed power of 98% (N = 13). To reach at least 80% power for an effect of this size required only five participants (power = 82%). The effect of condition was not reliable in the WebGazer data (β = -0.03, z = -0.18, p = 0.86; observed power 5% for N = 32). If we assume that the true effect in the WebGazer data was 25% the size of the effect in video data, then a sample of approximately 120 participants would be required to achieve power greater than 80% (power=84%). This sample is 24 times the required minimum for the video data effect size. This conjecture is based on the relative effect sizes in Experiment 1A, though it is of course possible that the true effect size for WebGazer is considerably larger, or smaller, than our estimate.

In sum, our results suggest that webcam video annotation is a far more sensitive means of detecting the kind of fine-grained eye-movement effects that are relevant to many child language researchers. WebGazer estimation may be better suited to detecting fairly long-lasting effects in which the primary outcome measure is which part of the screen participants fixated on.

4.4.3. Recommendations for practice and directions for future research

Our results suggest that while both webcam video annotation and WebGazer estimation can be used with child participants in web-based tasks, the two methods have different advantages and disadvantages.

Webcam video annotation has better spatiotemporal accuracy than WebGazer (drastically reducing the amount of noise in our child data), making the method better suited to detecting the temporally-sensitive, fine-grained looking patterns assessed in studies of real-time language processing. Collecting webcam video data over Zoom requires relatively little technical expertise, as the experiment itself can be built and run in any software; the experiment can either be run on the participant's computer (as in Experiment 1) or displayed from the experimenter's computer using Zoom's screen sharing function (as in unpublished work by Anthony Yacovone, personal communication). It is possible to collect webcam video for gaze annotation in unsupervised web experiments using Zoom (Slim et al., 2022) or other webcam recording functions (e.g., via PCIbex; Ovans, 2022). However, the hand annotation process is time consuming (in Experiment 1, annotating a seven second video took approximately one minute), and the resulting gaze location estimates are relatively coarse-grained (representing regions of the display instead of coordinate estimates).

WebGazer's gaze coding, on the other hand, is automatic, reducing the data processing burden on the researcher. It can be used to obtain either gaze coordinate estimates or binary looks to relevant screen locations, and the data are saved in text format, thus helping to maintain participant privacy and requiring less storage space than video recordings. Our results suggest that it is possible to achieve similar target look resolution with WebGazer in quadrant-based analyses in both supervised and unsupervised web-based studies. In addition, WebGazer is free to use and has implementations in popular frameworks for web-based research. However, use of these implementations often requires working proficiency in programming languages, and implementations may not be compatible with all web-browsers. Furthermore, WebGazer's low spatiotemporal accuracy makes it more difficult to detect fine-grained effects with sufficient resolution and power. The sample sizes required to detect such effects with sufficient power are much larger than for webcam video annotation (10x the size or greater). These sample sizes may be prohibitively large for experiments targeting smaller effects. Nevertheless, WebGazer was able to detect looks towards targets in both of our experiments, suggesting that it is suitable for tasks that require spatial discrimination of robust looking patterns.

It is possible that the quality of data collected with WebGazer would be improved by having participants complete the experiment in the same environment or with the same computer (e.g., Özgoy et al., 2023; Semmelmann & Weigelt, 2018) — for instance, if a researcher uses a laptop as a mobile lab. However, recent work in our lab with Mieke Slim and Anthony Yacovone suggests that the limitations of WebGazer persist under more controlled conditions; in a comparison of an infrared eye-tracker, WebGazer, and webcam video annotation, we found no substantial differences in eye-movement effects when the two webcam methods were applied in

the lab or in a web-based setting (where participants completed the experiment from their own computers).

Researchers should consider these trade-offs when deciding whether to conduct eye-tracking studies online and which gaze estimation method to use. For researchers interested in using WebGazer for online studies, we have several recommendations:

1) When designing the task, do not rely on vertical distinctions between critical stimuli. Consider simplifying the task to involve a two-image display or place critical stimuli on different halves of the screen in quadrant-based designs. Looks to diagonally-adjacent stimuli may be most easily discriminated (see Experiment 2).

2) When determining sample size, assume a 50–75% reduction in effect size relative to in-lab effects. Specifically, we found that the effects observed using WebGazer were roughly 25% as large (for the Experiment 1A cohort effect) to 45% as large (for horizontal target-side looks in the Experiment 1 control trials) as in the webcam video data, which produced effects of roughly the same magnitude as prior in-lab studies. The estimated reduction in effect size for WebGazer appears to vary based on effect type (short-lived, small effects vs. long-lasting fixations). Future work should investigate the performance of webcam eye-tracking methods in detecting various types of effects in order to provide more accurate recommendations for estimating expected effect sizes.

3) When planning the analysis, consider the likelihood of temporal delays in effect emergence. To account for such delays, researchers should shift or widen their planned analysis window appropriately or use an analysis method that does not assume a precise effect time window (e.g., cluster permutation analyses).

4) Consider setting calibration thresholds and/or including recalibration checkpoints to encourage participants to remain in an optimal position for WebGazer. Based on the data in Figure 4.14, we tentatively recommend a calibration threshold of at least 30% (though thresholds may need to be higher for smaller effects and/or more complicated displays).[12] To help improve WebGazer performance, ask parents to adjust their child's distance from the computer, the camera angle, and room lighting as necessary so that the participant's eyes can easily be seen in the webcam video feed at the onset of the calibration sequence.

For researchers interested in using webcam video annotation, we have the following recommendations:

1) Consider placing critical stimuli on different halves of the screen. While hand annotators were better at distinguishing quadrant looks than WebGazer in Experiment 1, horizontal differences are still easier for annotators to discriminate.

2) Re-center participant gaze with a central fixation prior to the onset of the experimental stimuli; having the gaze begin in the center of the screen makes it easier to identify in which direction looks are launched.

3) If using Zoom to record webcam video, utilize the gallery view layout for the recording (as opposed to the active speaker view) and hide non-video participants. This will ensure that the participant's webcam stream is present in the recording throughout the entire duration of the experiment. If the experimenter(s) turn off their video after starting the recording, the participant's face will be the only recorded view (note that at the time of writing, turning the

[12] See Supplementary Materials for an analysis of the cohort effect in the Experiment 1 WebGazer data restricted to trials that meet this calibration threshold. In Experiment 1A, the cohort effect cluster increased in size (800–1199 ms after target onset; z-sum = 9.42, $p < 0.01$) relative to our original analysis; there was still no cohort effect identified in the Experiment 1B WebGazer data. See Supplementary Materials for additional analyses relating trial calibration score to the size of the phonemic cohort effect in Experiments 1A and 1B.

experimenter video off prior to starting the recording causes Zoom to default to recording active speaker view). Zoom allows for simultaneous recordings in multiple layouts (e.g., screen recording, screen recording + thumbnail speaker view), which may be useful for aligning the recorded gaze data to trial onsets in the experiment.

4) If using teleconferencing software like Zoom to collect screen and/or webcam video recordings, test the available functions and settings for recordings. Some functions (e.g., Zoom's optimize for video function) may produce unexpected delays in the audio–visual sync within recordings.

5) Consider using a combination of visual and auditory prompts to identify trial onsets within the Zoom recordings; this will allow researchers to recover trial onsets should there be any issues with audio–visual synchrony in the recording.

6) Make sure that participant and/or experimenter has a way to view the participant webcam stream prior to the start of the experiment (e.g., through Zoom teleconference or by showing a video preview in PCIbex) so that the participant can adjust positioning and lighting to ensure that their eyes are visible in video recording.

While this work provides a starting point for evaluating online eye-tracking research with children, much remains to be done. For instance, future work should compare webcam-based eye-tracking methods to traditional high-end in-lab eye-trackers and should further assess the feasibility of running unsupervised web-based experiments with children. Despite the success of the Experiment 2 fixation task, we know very little about the limits of unsupervised tasks, particularly those with more complicated designs. In addition, while we observed age-related differences in WebGazer accuracy, it is unclear what is driving those differences, the extent to which they might influence the detection of linguistic effects, and whether we should expect

similar differences in annotated webcam videos. Finally, as improvements continue to be made to automatic gaze-coding algorithms, their performance with child populations will need to be re-assessed.

4.5. Conclusion

We have demonstrated in two experiments that it is possible to run web-based visual-world studies with school-aged children in both supervised and unsupervised experimental settings. We tested two webcam eye-tracking methods and found that they are differentially suitable for detecting different kinds of effects. While both methods can discriminate looks to a target (albeit with different levels of accuracy), we found that WebGazer is not well-suited to detecting effects that require a high level of spatiotemporal accuracy (see Slim & Hartsuiker, 2022 for a similar conclusion). In contrast, frame-by-frame annotation of gaze direction from webcam videos provided sufficient spatial and temporal resolution to detect a fleeting and subtle effect typical of those studied by child language researchers. We anticipate that webcam eye-tracking will continue to improve as researchers develop tools, experimental protocols, and practices that are more precise, accurate, and efficient. We hope that these improvements will allow child language researchers to take advantage of the benefits of large-scale web-based experimentation for eye-tracking research.

Chapter 5

[Conclusion]

5.1. Conclusion

During language processing, information moves across levels of representation in an incremental, cascading, and interactive fashion. Understanding when and how these features develop can shed light on their source, revealing whether the dynamics of language processing reflect basic, fundamental properties of the language system (consequences of the system architecture) or whether they emerge later as a product of domain-general cognitive efficiency or domain-specific experience. In this dissertation, I presented three papers that explore whether incremental interactive cascades are already present in the language system by early childhood, as well as the methods that can be used to ask this question, using lexical processing as a case study. This Conclusion provides a brief summary of the findings of each paper as well as the take-aways for our understanding of information flow in the human language system.

5.2. Summary of key findings

5.2.1. Paper 1: Cascaded processing in word production

Paper 1 explored whether cascaded processing is present in the developing language production system by five years of age. Specifically, the paper targeted how information flows

between lexical selection and phonological encoding processes during word production. In theory, information can pass through multilevel systems in one of two possible fashions: One possibility is that information moves through the system in a strictly serial fashion, with information only moving from one level to the next once processing at the first level is complete. In the case of lexical processing, this would mean that speakers do not begin accessing the phonological forms of words to be produced during phonological encoding until they have already selected which word of the available candidates to produce during lexical selection. Alternatively, information can cascade across the levels of the system, with later processes beginning before earlier ones are complete. In lexical processing, this would mean that activation spreads to the phonological forms of candidate words even before the speaker has selected which one to produce.

There is compelling evidence from speech error analyses and reaction time studies that information moves through the mature language production system in a cascaded fashion rather than a serial one (e.g., Costa et al., 2000; Cutting & Ferreira, 1999; Jescheniak & Schriefers, 1998; Morsella & Miozzo, 2002; Peterson & Savoy, 1998; Rapp & Goldrick, 2000; Starreveld & La Heij, 1995; inter alia). Paper 1 asked whether this cascaded processing is a fundamental product of the mind's architecture that arises early in life. Although there is clear (but limited) evidence for cascaded activation in word planning at seven years of age (Jescheniak et al., 2006), evidence for cascaded processing before this age is sparse and open to multiple interpretations. One potential reason why the dynamics of lexical processing remains underexplored from a developmental perspective is that the paradigms used to tap into these processes with adults are often not suitable for research with young children. For instance, a large body of prior adult work investigating cascading activation has relied on interference paradigms in which participants

must name an image in the presence of a simultaneously-presented distractor stimulus (e.g., Abdel Rahman & Melinger, 2008; Damian & Bowers, 2003; Damian & Martin, 1999; Humphreys et al., 2010; Jescheniak et al., 2005; Jescheniak et al., 2006; Jescheniak & Schriefers, 1998; Kuipers & La Heij, 2009; La Heij et al., 1990; Mädebach et al., 2011; Melinger & Abdel Rahman, 2013; Meyer & Damian, 2007; Morsella & Miozzo, 2002; Navarette et al., 2017; Navarrete & Costa, 2009; Roelofs, 1992; Roelofs, 2008; Schriefers et al., 1990; Vigliocco et al., 2004; Zhang et al., 2018; inter alia; see Chapter 2 for detail). Given age-related differences in children's responses to the cognitive/perceptual load introduced by interference paradigms (Jerger et al., 2013), these tasks may be more difficult for children below the age of six to seven years. Paper 1 introduced a novel way of testing for informational cascades in lexical processing, using a simple picture naming paradigm to explore a previously unstudied prediction of cascaded processing: that phonological activation begins while lexical selection is underway, resulting in interactions between variables that influence each process.

In a picture naming study with adults and five-year-old children, I manipulated picture codability (name agreement; e.g., Snodgrass & Vanderwart, 1980) and name frequency, factors that respectively affect the processes of lexical selection (Alario et al., 2004; Balatsou et al., 2022; Griffin, 2001; Johnson, 1992; inter alia) and phonological encoding (Jescheniak & Levelt, 1994; Griffin & Bock, 1998; though see Chapter 2 for discussion of frequency effects at the lexical/conceptual level). I investigated the influence of these two factors on response time in both populations. I replicated prior results showing that adults and children are faster to name pictures with higher codability (e.g., Butterfield & Butterfield, 1977; Johnson, 1992; Johnson & Clark, 1988; Lachman, 1973; Lachman et al, 1974; Lachman & Lachman, 1980; Paivio et al., 1989; inter alia) and whose names are more frequent (e.g., Bates et al., 2003; D'Amico et al.,

2001; Jescheniak & Levelt, 1994; Lachman, 1973; Lachman et al., 1974; Oldfield & Wingfield, 1965; inter alia). Ex-Gaussian distribution analyses suggested that the effects of codability and frequency had qualitatively different influences on the response time distribution from each other (supporting the hypothesis that they play different roles in lexical processing) but similar influences across participant populations, suggesting that similar processes underlie word production in adults and five-year-old children. In particular, the codability effect had a particularly pronounced skewing effect on the adult and child RT distributions, as would be expected if codability effects reflect increased time to resolve competition during lexical selection (skewing effects are a hallmark of decision-making processes; Hohle, 1965).

Critically, in addition to demonstrating independent effects of codability and frequency, the data from both populations displayed under-additive interactions between these effects such that the effect of frequency was attenuated when codability (i.e., name agreement) was low. To my knowledge, such an interaction has not been previously reported for either population (importantly, this is because studies typically explore codability and frequency effects separately, not because previous research has looked for an interaction and failed to find it). I observed comparable interactions in a secondary analysis of multilingual naming data from Bates et al. (2003), suggesting that this interaction effect generalizes across languages and stimulus sets. I confirmed via simulations that this form of interaction arises as a natural consequence of a cascading activation architecture.

Paper 1 thus provides evidence that the child language production system displays the same dynamic information flow as the adult system, with informational cascades robustly present in the system by five years of age. While the present study does not resolve the question of how and when cascaded processing develops, it places constraints on the answer, suggesting

that the capability is present early, as one would expect if such cascades are a fundamental property of the language system. Furthermore, the novel way of assessing cascaded processing introduced in Paper 1 will allow future research to explore the dynamics of information flow during lexical processing at younger ages than before and in other populations for whom complex task designs may be more difficult (e.g., individuals with neurodevelopmental disorders).

5.2.2. Paper 2: Interactive processing in word comprehension

Paper 1 provides evidence that informational cascades are present in the language production system by at least five years of age, with activation spreading to phonological forms before lexical selection is complete. There is also evidence for comparable informational cascades in the child language comprehension at around the same age, with activation spreading to semantic representations before phonological processing is complete (Huang & Snedeker, 2011). We thus see that information cascades through the language system as young children produce or comprehend words. An important follow-up question is whether information only cascades through the developing language system in a single direction (top-down for language production, bottom-up for language comprehension) or whether information can move in both top-down and bottom-up directions, as it does in the adult system. Paper 2 addresses this question by asking whether four and five-year-old children are able to use top-down information during language comprehension to guide word recognition processes and predict upcoming words.

In a visual world eye-tracking study with both adults and four and five-year-old children, participants viewed image displays and heard sentences that were either constraining towards a

target word (e.g., *The baby drank the **milk**…*) or neutral (e.g., *The child took the **milk**…*). In some trials, an image on screen had a name that started with the same sounds as the target word (e.g., ***mi**tten*). In the neutral sentences, participants looked at this cohort competitor as they heard the beginning of the target word, suggesting that it was initially considered as a potential match to the input because it started with the same sounds (replicating the phonemic cohort effect observed in prior work with neutral contexts; e.g., Allopenna et al., 1998; Dahan & Gaskell, 2007; Dahan et al., 2001; Desroches et al., 2006; Farris-Trimble & McMurray, 2013; Magnusson et al., 1999; Sekerina & Brooks, 2007; Rigler et al., 2015; Weighall et al., 2017; inter alia). In the constraining sentences, participants did not look at cohort competitors more than controls when hearing the target word, indicating that they were able to use top-down contextual information to rule out the competitor word, even though it initially matched the perceptual input. Critically, the target word (e.g., *milk*) was not included in the experimental displays (cf. Dahan & Tanenhaus, 2004), meaning that the lack of a phonemic cohort effect in the constraining sentences cannot be attributed to anticipatory looks to the target (e.g., Altmann & Kamide, 1999; Mani & Huettig, 2012) drawing looks away from the cohort competitor (see also Gaston, 2020 for additional discussion of the benefits of not including a target image when investigating competitor activation levels). These results illustrate that participants were able to use top-down contextual information to constrain the processing of the bottom-up input and avoid competition from semantically-incongruent competitors. There was no difference in this effect between age groups, suggesting that adult-like pathways for interactive processing are robust and active early in life.

Moreover, children's eye-movements in the constraining sentences showed evidence of phonological form prediction. Children looked to cohort competitor images shortly after target

words became predictable in the constraining sentences but well before the target words were articulated, suggesting that they were able to use top-down information to predict target words and activate their phonological forms (a form of predictive phonological competition effect; Ito, 2024). By looking for early looks to cohort competitor images instead of target images, the results of Paper 2 are able to go beyond prior work investigating prediction in young children's language comprehension, which has primarily targeted *referential prediction* (e.g., Borovsky et al., 2012; Borovsky et al., 2014; Lukyanenko & Fisher, 2016; Mani & Huettig, 2012; Nation et al., 2003; Özge et al., 2019; Sommerfeld et al., 2023; inter alia), in order to show that what is predicted can be specifically lexical in nature.

The findings of Paper 2 thus illustrate that top-down informational cascades and top-down and bottom-up interactivity are available during language comprehension at least by four to five years of age, despite prior findings showing reduced or no integration of top-down information during children's language comprehension at this same range and older (e.g., Joseph et al., 2008; Khanna & Boland, 2010; Snedeker & Trueswell, 2004; Snedeker & Yuan, 2008; Tiffin-Richards & Schroeder, 2020; Trueswell et al., 1999; Yacovone et al., 2021). As discussed in Paper 2, this contrast may result from differences in task demands, the available top-down cues, or type of processing under investigation (e.g., lexical processing vs. syntactic processing). The results of Paper 2 suggest that top-down and bottom-up interactive pathways develop early, which is relevant not only to our understanding of processing dynamics and their developmental origins within the language system but also to models of cognitive development more generally, as interactivity is also thought to be present in many other cognitive processes, including visual perception (e.g., Bullier, 2001; Hochstein & Ahissar, 2002; Kafaligonul et al., 2015), olfactory perception (e.g., Andersson et al., 2018), gustatory perception (e.g., Kobayashi et al., 2004),

somatosensory perception (e.g., Haegens et al., 2011), object recognition (e.g., Fenske et al., 2006; Panichello et al., 2012; Wyatte et al., 2014), and face perception (e.g., Rossion et al., 2003; Sorger et al., 2007).

5.2.1. Paper 3: Using webcam eye-tracking to assess real-time language processing

Paper 3 takes a closer look at the visual world paradigm used in Paper 2 and explores whether it is possible to move this paradigm to web-based settings by using webcam eye-tracking techniques. In this study, I compared two methods of webcam eye-tracking for use in child language research: automatic gaze estimation from the webcam video feed using WebGazer.js (Papoutsaki et al., 2016) and frame-by-frame annotations of gaze direction from webcam videos recorded during the experiment session (e.g., Ovans, 2022; Slim, Kandel et al. 2024; see Huang & Snedeker, 2013; Snedeker & Trueswell, 2004; Thothathiri & Snedeker, 2008; Yacovone et al., 2021 for lab-based applications).

The study comprised two experiments. The first experiment compared these two methods in a study of lexical activation with five and six-year-old children, testing how well the methods discriminated both robust fixation patterns (looks to target stimuli) as well as more subtle eye-movement patterns of the kind relevant to child language researchers — in particular, phonemic cohort competition effects, which serve as evidence for incrementality in real-time language comprehension (e.g. Allopenna et al., 1998; Sekerina & Brooks, 2007). Data collection for the two methods occurred simultaneously for the same participants, allowing for a direct comparison of the methods' performance. The second experiment assessed WebGazer's performance with children aged 4–12 years of age in a simple visual fixation task.

The results of Paper 3 demonstrate that it is indeed possible to run web-based eye-tracking experiments with young children, however the two webcam-based eye-tracking methods differed in their sensitivity and accuracy. Webcam video annotation was well-suited to detecting both robust fixation patterns as well fine-grained, spatiotemporally sensitive real-time processing effects, such as the phonemic cohort effect (providing additional evidence for incrementality in young children's language comprehension; e.g., Desroches et al., 2006; Sekerina & Brooks, 2007; Rigler et al., 2015; Weighall et al., 2017; inter alia). In contrast, WebGazer.js provided estimates that were noisier and less precise, making the method best-suited for detecting effects that require simple spatial discrimination of large, stable looking patterns. Paper 3 offers recommendations for conducting web-based eye-tracking studies with children, which are applicable for both child language research and developmental psychology research more generally. The results of Paper 3 suggest that, as webcam eye-tracking technology continues to evolve, webcam video annotation can serve as a useful benchmark against which to compare automatic gaze estimation algorithms for research with children (see also Slim, Kandel et al., 2024 for webcam video annotation with adults and for comparison to in-lab infrared eye-tracking). WebGazer's apparent reduced performance with children relative to adults (Slim & Hartsuiker, 2022; Slim, Kandel et al., 2024) further highlights the importance of considering population-specific characteristics when developing and applying gaze estimation algorithms.

5.3. Concluding summary

In sum, the projects in this dissertation demonstrate that the incremental, interactive informational cascades that are integral to adult language processing are robust and active in the language processing of young children. Paper 1 provides evidence for informational cascades in

language production in five-year-old children, and Paper 2 provides evidence for interactive cascades in language comprehension in four and five-year-olds. These results place constraints on our theories of when and how these capacities emerge, showing that they are present in early childhood, which could suggest that they are not simply learned strategies but rather intrinsic properties of the language system. In addition to the theoretical contributions made by the present work regarding the development of processing dynamics in the human language system, the research in this dissertation also makes methodological contributions to the study of lexical processing (and real-time processing more generally) in young children — introducing a new way to assess informational cascades in language production (Paper 1), teasing apart what paradigms can be used to assess referential versus lexical prediction (Paper 2), and providing guidance for researchers interested in running web-based eye-tracking studies with young children (Paper 3).

In sum, the present work illustrates that the information dynamics governing adult language processing are present from early childhood, supporting the hypothesis that incremental, cascaded, and interactive processing dynamics are fundamental properties of the human language system. However, there is still much work to be done. Future work should continue to explore the dynamics of information flow in the developing language system, testing at what age these key properties first arise, whether they develop on similar timelines for different forms of language processing (see, e.g., Chapter 3 for discussion of potential differences between lexical and syntactic processing in language comprehension), how these properties evolve and change from their initial onset, and the consequences of these properties for theories of language acquisition and later language development, such as literacy acquisition. By continuing to investigate these questions and others, we can deepen our understanding of the

mechanisms and cognitive structures that underlie human language processing and development, as well as the human mind more generally. In addition, understanding the principles of information processing in the language system and how they develop has the potential to inform educational practices, guide language interventions, and lead to innovations in modelling human-like processing and language acquisition in artificial intelligence.

References

Abdel Rahman, R., & Melinger, A. (2008). Enhanced phonological facilitation and traces of concurrent word form activation in speech production: An object-naming study with multiple distractors. *Quarterly Journal of Experimental Psychology, 61*(9), 1410–1440. https://doi.org/10.1080/17470210701560724

Adamson-Harris, A. (2000). *Processing semantic and grammatical information in auditory sentences: electrophysiological evidence from children and adults.* [Unpublished doctoral dissertation]. University of Oregon.

Alameda, J. R., & Cuetos, F. (1995). *Diccionario de Frecuencias de las Unidades Lingüísticas del Castellano* [Frequency dictionary for lexical items in Castilian Spanish]. Servicio de Publicaciones de la Universidad de Oviedo.

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers, 36*(1), 140–155. https://doi.org/10.3758/BF03195559

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., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*(3), 191–238. https://doi.org/10.1016/0010-0277(88)90020-0

Andersen, E. S. (2014). *Speaking with Style: The Sociolinguistic Skills of Children (RLE Linguistics C: Applied Linguistics)*. Routledge. https://doi.org/10.4324/9781315856902

Anderson, J. D. (2008). Age of acquisition and repetition priming effects on picture naming of children who do and do not stutter. *Journal of Fluency Disorders, 33*(2), 135–155. https://doi.org/10.1016/j.j udis.2008.04.001

Andersson, L., Sandberg, P., Olofsson, J. K., & Nordin, S. (2018). Effects of task demands on olfactory, auditory, and visual event-related potentials suggest similar top-down modulation across senses. *Chemical Senses, 43*(2), 129–134. https://doi.org/10.1093/chemse/bjx082

Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(2), 514–544. https://doi.org/10.1037/0278-7393.27.2.514

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkam, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavioral Research Methods, 52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Atkinson, E., Wagers, M. W., Lidz, J., Phillips, C., & Omaki, A. Developing incrementality in filler-gap dependency processing. *Cognition, 179*, 132–149. https://doi.org/10.1016/j.cognition.2018.05.022

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 14*(4), 382–391. https://doi.org/10.1016/S0022-5371(75)80017-X

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database* (Release 2) [CD-ROM]. Linguistic Data Consortium.

Balatsou, E., Fischer-Baum, S., & Oppenheim, G. M. (2022). The psychological reality of picture name agreement. *Cognition, 218*, Article 104947. https://doi.org/10.1016/j.cognition.2021.104947

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128*(1), 32–55. https://doi.org/10.1037/0096-3445.128.1.32

Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language, 59*(4), 495–523. https://doi.org/10.1016/j.jml.2007.10.004

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C. C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G.,

Hung, D., & Hsu, J. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review, 10*(2), 344–380. https://doi.org/10.3758/BF03196494

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bentin, S., McCarthy, G., & Wood C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neuropsychology*, *60*(4), 343–355. https://doi.org/10.1016/0013-4694(85)90008-2

Bergelson, E. (2020). The comprehension boost in early word learning: Older infants are better learners. *Child Development Perspectives, 14*(3), 142–149. https://doi.org/10.1111/cdep.12373

Bergelson, E., & Aslin, R. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences, 114*(49), 12916–12921. https://doi.org/10.1073/ pnas.1712966114

Bergelson, E., & Swingley, D. (2012). At 6 to 9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253–3258. https://doi.org/10.1073/ pnas.1113380109

Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young infants. *Cognition, 127*(3), 391–397. https:// doi.org/10.1016/j.cognition.2013.02.011

Bergelson, E., & Swingley, D. (2015). Early word comprehension in infants: Replication and extension. *Language Learning and Development, 11*(4), 369–380. https://doi.org/10.1080/ 15475441.2014.979387

Best, J. R. & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development, 81*(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Bjorklund, D. F. (1995). *Children's Thinking.* Brooks/Cole.

Bjorklund, D. F., & Harnishfeger, K. K. (1990). The resources construct in cognitive development: Diverse sources of evidence and a theory of inefficient inhibition. *Developmental Review, 10*(1), 48–71. https://doi.org/10.1016/0273-2297(90)90004-N

Blanken, G. (1998). Lexicalisation in speech production: Evidence from form-related word substitutions in aphasia. *Cognitive Neuropsychology, 15*(4), 321–360. https://doi.org/10.1080/026432998381122

Bloem, I., & La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and Language, 48*(3), 468–488. https://doi.org/10.1016/S0749-596X(02)00503-X

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, *89*(1), 1–47. https://doi.org/10.1037/0033-295X.89.1.1

Bock, J. K. (1995). Sentence production: From mind to mouth. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of Perception and Cognition. Vol. 11: Speech, Language, and Communication* (pp. 181–216). Academic Press. https://doi.org/10.1016/B978-012497770-9.50008-X

Bock, J. K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). Academic Press.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341–345.

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology, 112*(4), 417–436. https://doi.org/10.1016/j.jecp.2012.01.005

Borovsky, A., Sweeney, K., Elman, J. L., & Fernald, A. (2014). Real-time interpretation of novel events across childhood. *Journal of Memory and Language, 73*, 1–14. https://doi.org/10.1016/j.jml.2014.02.001

Brothers, T., Dave, S., Hoversten, L. J., Traxler, M. J., & Swaab, T. Y. (2019). Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes. *Neuropsychologia, 135*, Article 107225. https://doi.org/10.1016/j.neuropsychologia.2019.107225

Brouwer, S., Özkan, D., & Küntay, A. (2019). Verb-based prediction during language processing: The case of Dutch and Turkish. *Journal of Child Language, 46*(1), 80–97. https://doi.org/10.1017/S0305000918000375

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*(1), 34-44. https://doi.org/10.1162/jocn.1993.5.1.34

Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2002). Reference Resolution in the Wild: On-line circumscription of referential domains in a natural, interactive Problem-solving task. *Proceedings of the Annual Meeting of the Cognitive Science Society, 24*.

Brysbaert, M., & Ghyselinck, M. (2007). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition, 13*(7–8), 992–1011. https://doi.org/10.1080/13506280544000165

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

Budd, M.-J., Hanley, R., & Griffiths, Y. (2011). Simulating children's retrieval errors in picture-naming: A test of Foygel and Dell's (2000) semantic/phonological model of speech production. *Journal of Memory and Language, 64*(1), 74–87. https://doi.org/10.1016/j.jml.2010.08.005

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews, 36*(2–3), 96–107. https://doi.org/10.1016/S0165-0173(01)00085-6

Butterfield, G. B., & Butterfield, E. C. (1977). Lexical codability and age. *Journal of Verbal Learning and Verbal Behavior, 16*(1), 113–118. https://doi.org/10.1016/S0022-5371(77)80013-3

Butterworth, B. (1981). Speech errors: Old data in search of new theories. *Linguistics, 19*(7-8), 627–662. https://doi.org/10.1515/ling.1981.19.7-8.627

Butterworth, B. (1989). Lexical access in speech production. In W. Marslen-Wilson (Ed.), *Lexical Representation and Process* (pp. 108– 135). Cambridge, MA: MIT Press.

Butterworth, B. (1992). Disorders of phonological encoding. *Cognition, 42*(1-3), 261–286. https://doi.org/10.1016/0010-0277(92)90045-J

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*(1), 177–208. https://doi.org/10.1080/026432997381664

Carroll, J. B., & White, M. N. (1973). Word frequency and age-of-acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology, 25*(1), 85–95. https://doi.org/10.1080/14640747308400325

Carter, C. S., Mintun, M., & Cohen, J. D. (1995). Interference and facilitation effects during selective attention: An H2150 PET study of Stroop task performance. *NeuroImage, 2*(4), 264–272. https://doi.org/10.1006/nimg.1995.1034

Carver A. C., Livesey D. J., & Charles M. (2001). Age related changes in inhibitory control as measured by stop signal task performance. *International Journal of Neuroscience*, *107*(1-2), 43–61. https://doi.org/10.3109/00207450109149756

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. S. Siegler (Ed.), *Children's Thinking: What Develops?* (pp. 73–96). Lawrence Erlbaum.

Chinese Knowledge Information Processing Group. (1997). *Academia sinica balanced corpus (WWW Version 3.0)*. Institute of Information Science, Academia Sinica.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (pp. 127–149). American Psychological Association. https://doi.org/10.1037/10096-006

Collins, A. M., & Loftus, E. F. (1975). Spreading activation theory of semantic processing. *Psychological Review, 82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Comalli Jr., P. E., Wapner, S., & Werner, H. (1962). Interference effects of Stroop colour-word test in childhood, adulthood, and aging. *The Journal of Genetic Psychology, 100*(1), 47–53. https://doi.org/10.1080/00221325.1962.10533572

Contemori, C., Carlson, M., & Marnis, T. (2018). On-line processing of English which-questions by children and adults: A visual world paradigm study. *Journal of Child Language, 45*(2), 415–441. https://doi.org/10.1017/S0305000917000277

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

Cooper-Cunningham, R., Charest, M., Porretta, V., & Järvikivi, J. (2020). When couches have eyes: The effect of visual context on children's reference processing. *Frontiers in Communication, 5*, Article 576236. http://doi.org/10.3389/fcomm.2020.576236

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1283–1296. https://doi.org/ 10.1037/0278-7393.26.5.1283

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review, 24*(4), 1158–1170. https://doi.org/10.3758/s13423-016-1191-6

Cowan, N., Fristoe, N. M., Elliott, E. M., Brunner, R. P., & Saults, J. S. (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory & Cognition, 34*(8), 1754–1768. https://doi.org/10.3758/BF03195936

Cunningham, A. E. & Stanovich, K. E. (1998). What reading does for the mind. *American Educator, 22*(1-2), 8–15.

Cutler, A. & Clifton, C. E. (1999). Comprehending spoken language: A blueprint of the listener. In C. Brown & P. Hagoort (Eds.), *Neurocognition of Language* (pp. 123–166). Oxford University Press.

Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 318–344. https://doi.org/10.1037/0278-7393.25.2.318

Cycowicz, Y. M., Friedman, D., & Rothstein, M. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology, 65*(2), 171–237. https://doi.org/10.1006/jecp.1996.2356

Dahan, D., & Gaskell, M. G. (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language, 57*(4), 483–501. https://doi.org/10.1016/j.jml.2007.01.001

Dahan, D., Magnusson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317–367. https://doi.org/10.1006/cogp.2001.0750

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language, 42*(4), 465–480. https://doi.org/10.1006/jmla.1999.2688

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 498–513. https://doi.org/10.1037/0278-7393.30.2.498

Dalrymple, K. A., Manner, M. D., Harmelink, K. A., Teska, E. P., & Elison, J. T. (2018). An examination of recording accuracy and precision from eye tracking data from toddlerhood to adulthood. *Frontiers in Psychology, 9*, Article 803, https://doi.org/10.3389/fpsyg.2018.00803

Damian, M. F., & Bowers, J. S. (2003). Locus of semantic interference in picture-word interference tasks. *Psychonomic Bulletin & Review, 10*(1), 111–117. https://doi.org/10.3758/BF03196474

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 345–361. https://doi.org/10.1037/0278-7393.25.2.345

D'Amico, S., Devescovi, A., & Bates, E. (2001). Picture naming and lexical access in Italian children and adults. *Journal of Cognition and Development, 2*(1), 71–105. https://doi.org/10.1207/S15327647JCD0201_4

Dawson, M. R. (1988). Fitting the ex-Gaussian equation to reaction time distributions. *Behavior Research Methods, Instruments, & Computers, 20*(1), 54–57. https://doi.org/10.3758/BF03202603

Deese, J. (1984). *Thought into Speech: The Psychology of a Language.* Prentice-Hall. https://doi.org/10.2307/413465

Degen, J., Kursat, L., & Leigh, D. (2021). Seeing is believing: Testing an explicit linking assumption for visual world eye-tracking in psycholinguistics. *Proceedings of the Annual Meeting of the Cognitive Science Society, 43*(1), 1500–1506. https://escholarship.org/uc/item/6182t9jb

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 369*(1634), Article 20120394. https://doi.org/10.1098/rstb.2012.0394

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review, 98*(4), 604– 614. https://doi.org/10.1037/0033-295X.98.4.604

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42*(1–3), 287–314. https:// doi.org/10.1016/0010-0277(92)90046-K

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20*(6), 611–629. https://doi.org/10.1016/S0022-5371(81)90202-4

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*(4), 801–838. https://doi.org/ 10.1037/0033-295X.104.4.801

DeLong, K. A., Chan, W.-H., & Kutas, M. (2019). Similar time courses for word form and meaning preactivation during sentence comprehension. Psychophysiology, 56(4), e13312. https://doi.org/10.1111/psyp.13312

DeLong, K. A., Chan, W.-H., & Kutas, M. (2020). Testing limits: ERP evidence for word form preactivation during speeded sentence reading. Psychophysiology, 58(2). https://doi.org/10.1111/psyp.13720

De Mauro, T., Mancini, F., Vedovelli, M., & Voghera, M. (1993). *Lessico di Frequenza dell'Italiano Parlato* [Frequency lexicon of spoken Italian]. ETASLIBRI.

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89*(1), 63–100. https://doi.org/10.1037/0033-2909.89.1.63

Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review, 12*(1), 45–75. https://doi.org/10.1016/0273-2297(92)90003-K

Desroches, A. S., Joanisse, M. F., & Robertson, E. K. (2006). Specific phonological impairments in dyslexia revealed by eyetracking. *Cognition, 100*(3), B32–B42. https://doi.org/10.1016/j.cognition.2005.09.001

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology, 71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Ehinger, B. V., Groß, K., Ibs, I., & König, P. (2019). A new comprehensive eye-tracking test battery concurrently evaluating the Pupil Labs glasses and the EyeLink 1000. *PeerJ, 7*, e7086. https://doi.org/10.7717/peerj.7086

Erel, Y., Shannon, K. A., Chu, J., Scott, K., Kline Struhl, M., Cao, P., Tan, X., Hart, P., Raz, G., Piccolo, S., Mei, C., Potter, C., Jaffe-Dax, S., Lew-Williams, C., Tenenbaum, J., Fairchild, K., Barmano, A., & Liu, S. (2022, May 1). iCatcher+: Robust and automated annotation of infant's and young children's gaze direction from videos collected in laboratory, field, and online studies. *PsyArXiv.* https://doi.org/10.31234/osf.io/up97k

Farris-Trimble, A. & McMurray, B. (2013). Test-retest reliability of eye tracking in the visual world paradigm for the study of real-time spoken word recognition. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1328–1345. https://doi.org/10.1044/1092-4388(2012/12-0145)

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*(4), 469–495. https://doi.org/10.1006/jmla.1999.2660

Fenske, M. J., Aminoff, E., Gronau, N., Bar. M. (2006). Top-down facilitation of visual object recognition: Object-based and context-based contributions. *Progress in Brain Research, 155*(B), 3–21. https://doi.org/10.1016/S0079-6123(06)55001-0

Fernald, A., Marchman, V. A., & Hurtado, N. (2008). Input Affects Uptake: How Early Language Experience Influences Processing Efficiency and Vocabulary Learning. *IEEE International Conference on Development and Learning*, *7*, 37–42. Monterey, CA, USA, 2008, pp. 37-42. https://doi.org/10.1109/DEVLRN.2008.4640802

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42*(1), 98–116. https://doi.org/10.1037/0012-1649.42.1.98

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science, 9*(3), 228–231. https://doi.org/10.1111/1467-9280.00044

Fernald, A., Swingley, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development*, *72*(4), 1003–1015. https://doi.org/10.1111/1467-8624.00331

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*(3), 348–368. https://doi.org/10.1016/0749-596X(86)90006-9

Ferreira, V., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(6), 1187–1199. https://doi.org/10.1037/0278-7393.28.6.1187

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language, 46*(1), 57–84. https://doi.org/10.1006/jmla.2001.2797

Fields, E. C., & Kuperberg, G. R. (2020). Having your cake and eating it too: Flexibility and power with mass univariate statistics for ERP data. *Psychophysiology*, *57*(2), e13468. https://doi.org/10.1111/psyp.13468

Finocchiaro, C., & Caramazza, A. (2006). The production of pronominal clitics: Implications for theories of lexical access. *Language and Cognitive Processes, 21*(1-3), 141–180. https://doi.org/10.1080/01690960400001887

Fox, J., & Weisberg, S. (2018). Visualizing fit and lack of fit in complex regression models with predictor effect plots and partial residuals. *Journal of Statistical Software, 87*(9), 1–27.

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd ed.). https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language, 43*(2), 182–216. https://doi.org/10.1006/jmla.2000.2716

Fraser, A., Gattas, S. U., Hurman, K., Robison, M., Duta, M., & Scerif, G. (2021, June 22). Automated gaze direction scoring from videos collected online through conventional webcam. *PsyArXiv.* https://doi.org/10.31234/osf.io/4dmjk

Friedmann, N., Biran, M., & Dotan, D. (2013). Lexical retrieval and its breakdown in aphasia and developmental language impairment. In C. Boeckx & K. K. Grohmann (Eds.), *The Cambridge Handbook of Biolinguistics* (pp. 350–374). Cambridge University Press.

Füredi, M., & Kelemen, J. (1989). *A Mai Magyar Nyelv Szépprózaigyakorisgi Szótra 1965–1977* [Prose frequency dictionary of contemporary Hungarian language 1965–1977]. Akadémiai Kiadó.

Gambi, C., Gorrie, F., Pickering, M. J., & Rabagliati, H. (2018). The development of linguistic prediction: Predictions of sound and meaning in 2- to 5-year-olds. *Journal of Experimental Child Psychology*, *173*, 351–370. https://doi.org/10.1016/j.jecp.2018.04.012

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience, 8*(2), 89–106. https://doi.org/10.1162/jocn.1996.8.2.89

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37*(1), 58–93. https://doi.org/10.1006/jmla.1997.2512

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language Production*, Vol. 1 (pp. 177-220). Academic Press.

Gaston, P. (2020). *The role of syntactic prediction in auditory word recognition*. [Doctoral Dissertation, University of Maryland, College Park]. Digital Repository at the University of Maryland. https://doi.org/10.13016/19yo-htfb

Gaston, P., Lau, E., & Phillips, C. (2020, December 4). How does(n't) syntactic context guide auditory word recognition? *PsyArXiv.* https://doi.org/10.31234/osf.io/sbxpn

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511790942

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language, 35*(3), 515. https://doi.org/10.1017/S0305000907008641

Goulden, R., Nation, P., & Read, J. (1990). How large can receptive vocabulary be? *Applied Linguistics, 11*(4), 341–363. https://doi.org/10.1093/applin/11.4.341

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition, 82*(1), B1–B14. https://doi.org/10.1016/S0010-0277(01)00138-X

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review, 10*(3), 603–609. https://doi.org/10.3758/BF03196521

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word recognition. *Journal of Memory and Language, 38*(3), 313–338. https://doi.org/10.1006/jmla.1997.2547

Griffin, Z., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*(4), 274 279. https://doi.org/10.1111/1467-9280.00255

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*(12), 1711–1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Guttentag, R. E., & Haith, M. M. (1978). Automatic processing as a function of age and reading ability. *Child Development, 49*(3), 707–716. https://doi.org/10.2307/1128239

Haegens, S., Händel, B. F., & Jensen, O. (2011). Top-down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task. *Journal of Neuroscience, 31*(14), 5197–5204. https://doi.org/10.1523/JNEUROSCI.5199-10.2011

Hahn, N., Snedeker, J., & Rabagliati, H. (2015). Rapid linguistic ambiguity resolution in young children with autism spectrum disorder: Eye tracking evidence for the limits of weak central coherence. *Autism Research*, *8*(6), 717–726. https://doi.org/10.1002/aur.1487

Hale, S. (1990). A global developmental trend in cognitive processing speed. *Child Development, 61*(3), 653–663. https://doi.org/10.2307/1130951

Hardy, M. A. (1993). *Regression with Dummy Variables.* Sage Publications, Inc.

Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science, 8*(3), 191–219. https://doi.org/10.1207/s15516709cog0803_1

Harley, T. A. (1993). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes, 8*(3), 291–309. https://doi.org/10.1080/01690969308406957

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61–83. https://doi.org/10.1017/s0140525x0999152x

Hochstein, S. & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*(5), 791–804. https://doi.org/10.1016/s0896-6273(02)01091-7

Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology, 69*(4), 382–386. https://doi.org/10.1037/h0021740

Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language, 35*(1), 66–85. https://doi.org/10.1016/0093-934X(88)90101-0

Holler, J. & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes*, *24*(2)*,* 267–289. https://doi.org/10.1080/01690960802095545

Horton, W. S. & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*(1), 91–117. https://doi.org/10.1016/0010-0277(96)81418-1

Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology, 45*(6), 1723–1739. https://doi.org/10.1037/a0016704

Huang, Y. T., & Snedeker, J. (2011). Cascading activation across levels of representation in children's lexical processing. *Journal of Child Language*, *38*(3), 644–661. https://doi.org/10.1017/s0305000910000206

Huang, Y. T., & Snedeker, J. (2013). The use of lexical and referential cues in children's online interpretation of adjectives. *Developmental Psychology, 49*(6), 1090–1102. https://doi.org/10.1037/a0029477

Huettig, F., & Brouwer, S. (2015). Delayed anticipatory spoken language processing in adults with dyslexia—evidence from eye-tracking. *Dyslexia: An International Journal of Research and Practice, 21*(2), 97–122. https://doi.org/10.1002/dys.1497

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience, 31*(1), 80–93. https://doi.org/10.1080/23273798.2015.1047459

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language, 57*(4), 460–482. https://doi.org/10.1016/j.jml.2007.02.001

Huettig, F., & Pickering, M. J. (2019). Literacy advantages beyond reading: Prediction of spoken language. *Trends in Cognitive Sciences, 23*(6), 464–475. https://doi.org/10.1016/j.tics.2019.03.008

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Humphreys, K., Boyd, C., & Watter, S. (2010). Phonological facilitation from pictures in a word association task: Evidence for routine cascaded processing in spoken word production. *Quarterly Journal of Experimental Psychology, 63*(12), 2289–2296. https://doi.org/10.1080/17470218.2010.509802

Ito, A. (2024). Phonological prediction during comprehension: A review and meta-analysis of visual-world eye-tracking studies. *Journal of Memory and Language, 139*, Article 104553. https://doi.org/10.1016/j.jml.2024.104553

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language, 86*, 157–171. https://doi.org/10.1016/j.jml.2015.10.007

Ito, A., & Husband, E. M. (2017). How robust are effects of semantic and phonological prediction during language comprehension? A visual world eye-tracking study. *IEICE Technical Report, 117*(149), 1–6.

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language, 98,* 1–11. https://doi.org/10.1016/j.jml.2017.09.002

Ito, A., & Sakai, H. (2021). Everyday language exposure shapes prediction of specific words in listening comprehension: A visual world eye-tracking study. *Frontiers in Psychology, 12*, Article 607474. https://doi.org/10.3389/fpsyg.2021.607474

Jaeger, J. J. (2005). *Kid's Slips: What Young Children's Slips of the Tongue Reveal about Language.* Lawrence Erlbaum Associates.

James, A. N., Minnihan, C. J., Watson, D. G. (2023). Language Experience Predicts Eye Movements During Online Auditory Comprehension. *Journal of Cognition, 6*(1), Article 30. https://doi.org/10.5334/joc.285

Jerger, S., Damian, M. F., Mills, C., Bartlett, J., Tye-Murray, N., & Abidi, H. (2013). Effects of perceptual load on semantic access by speech in children. *Journal of Speech, Language,*

and Hearing Research, 56(2), 388–403. https://doi.org/10.1044/1092-4388(2012/11-0186)

Jerger, S., Martin, R. C., & Damian, M. F. (2002). Semantic and phonological influences on picture naming by children and teenagers. *Journal of Memory and Language, 47*(2), 229–249. https://doi.org/10.1016/S0749-596X(02)00002-5

Jerger, S., Martin, R., & Pirozzolo, F. (1988). A developmental study of the auditory Stroop effect. *Brain and Language, 35*(1), 86–104. https://doi.org/10.1016/0093-934X(88)90102-2

Jerger, S., Pearson, D., & Spence, M. (1999). Developmental course of auditory processing interactions: Garner interference and Simon interference. *Journal of Experimental Child Psychology, 74*(1), 44–67. https://doi.org/10.1006/jecp.1999.2504

Jescheniak, J. D., Hahne, A., Hoffman, S., & Wagner, C. (2006). Phonological activation of category coordinates during speech planning is observable in children but not in adults: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(2), 373–386. https://doi.org/10.1037/0278-7393.32.3.373

Jescheniak, J. D., Hahne, A., & Schriefers, H. (2003). Information flow in the mental lexicon during speech planning: Evidence from event-related brain potentials. *Cognitive Brain Research, 15*(3), 261–276. https://doi.org/10.1016/S0926-6410(02)00198-2

Jescheniak, J. D., Hantsch, A., & Schriefers, H. (2005). Context effects on lexical choice and lexical activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 905–920. https://doi.org/10.1037/0278-7393.31.5.905

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 824–843. https://doi.org/10.1037/0278-7393.20.4.824

Jescheniak, J. D., & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(5), 1256–1274. https://doi.org/10.1037/0278-7393.24.5.1256

Jescheniak, J. D., & Schriefers, H. (2001). Priming effects from phonologically related distractors in picture–word interference. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 54A*(2), 371–382.

Johnson, C. J. (1992). Cognitive components of naming in children: Effects of referential uncertainty and stimulus realism. *Journal of Experimental Child Psychology, 53*(1), 24–44. https://doi.org/10.1016/S0022-0965(05)80003-7

Johnson, J. (2011). N400 Sentence-Level ERPs for 3, 4, and 5 Year Olds [Poster Presentation]. Annual Convention of the American Speech-Language-Hearing Association. San Diego, California, United States.

Johnson, C. J., & Clark, J. M. (1988). Children's picture naming difficulty and errors: Effects of age of acquisition, uncertainty, and name generality. *Applied Psycholinguistics, 9*(4), 351–365. https://doi.org/10.1017/S0142716400008055

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin, 120*(1), 113–139. https://doi.org/10.1037/0033-2909.120.1.113

Johnstone S. J., Pleffer C. B., Barry R. J., Clarke A. R., & Smith J. L. (2005). Development of inhibitory processing during the go/nogo task. *Journal of Psychophysiology, 19*(1), 11–23. https://doi.org/10.1027/0269-8803.19.1.11

Joseph, H. S., Liversedge, S. P., Blythe, H. I., White, S. J., Gathercole, S. E., & Rayner, K. (2008). Children's and adults' processing of anomaly and implausibility during reading: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*, *61*(5), 708–723. https://doi.org/10.1080/17470210701400657

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

Kafaligonul, H., Breitmeyer, B. G., & Öğmen, H. (2015). Feedforward and feedback processes in vision. *Frontiers in Psychology, 6,* Article 279. https://doi.org/10.3389/fpsyg.2015.00279

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin, 109*(3), 490–501. https://doi.org/10.1037/0033-2909.109.3.490

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica, 86*(2-3), 199–225. https://doi.org/10.1016/0001-6918(94)90003-5

Kampa, A., & Papafragou, A. (2020). Four-year-olds incorporate speaker knowledge into pragmatic inferences. *Developmental Science, 23*(3), Article e12920. https://doi.org/10.1111/desc.12920

Kandel, M., & Snedeker, J. (2024). Assessing two methods of webcam-based eye-tracking for child language research. *Journal of Child Language*, 1–34. https://doi.org/10.1017/S0305000924000175

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*(2), 201–258. https://doi.org/10.1207/s15516709cog1102_5

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition, 14*(2), 185–209. https://doi.org/10.1016/0010-0277(83)90029-X

Khanna, M. M., & Boland, J. E. (2010). Children's use of language context in lexical ambiguity resolution. *Quarterly Journal of Experimental Psychology*, *63*(1), 160–193. https://doi.org/10.1080/17470210902866664

Kidd, E., Stewart, A. J., & Serratrice, L. (2011). Children do not overcome lexical biases where adults do: The role of the referential scene in garden-path recovery. *Journal of Child Language, 38*(1), 222–234. https://doi.org/10.1017/S0305000909990316

Kim, A. E., & Lai, V. T. (2012). Rapid interactions between lexical semantic and word form analysis during word recognition in context: Evidence from ERPs. *Journal of Cognitive Neuroscience, 24*(5), 1104–1112. https://doi.org/10.1162/jocn_a_00148

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology, 25*(4), 463–492. https://doi.org/10.1080/02643290701674851

Kleiman, E. (2021). *EMAtools: Data Management Tools for Real-time Monitoring/Ecological Momentary Assessment Data.* R package version 0.1.4. https://CRAN.R-project.org/package=EMAtools

Kobayashi, M., Takeda, M., Hattori, N., Fukunaga, M., Sasabe, T., Inoue, N., Nagai, Y., Sawada, T., Sadato, N., & Watanabe, Y. (2004). Functional imaging of gustatory perception and imagery: "top-down" processing of gustatory signals. *NeuroImage, 23*(4), 1271–1282. https://doi.org/10.1016/j.neuroimage.2004.08.002

Krauss, R. M. & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4*(3), 343–346. https://doi.org/10.1037/h0023705

Kuipers, J.-R., & La Heij, W. (2009). The limitations of cascading in the speech production system. *Language and Cognitive Processes, 24*(1), 120–135. https://doi.org/10.1080/01690960802234177

Kukona, A. (2020). Lexical constraints on the prediction of form: Insights from the visual world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(11), 2153–2162. https://doi.org/10.1037/xlm0000935

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavioral Research Methods, 53*(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Kuperberg, G. R. & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, *8*(4), 533–572. https://doi.org/10.1080/01690969308407587

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146%2Fannurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161–163. https://doi.org/10.1038/307161a0

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). Lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lachman, R. (1973). Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology, 99*(2), 199–208. https://doi.org/10.1037/h0034633

Lachman, R., & Lachman, J. L. (1980). Picture naming: Retrieval and activation of long-term memory. In L. W. Poon, J. L. Fozard, L. S. Cermak, D. Arenberg, & L. W. Thompson (Eds.), *New Directions in Memory and Aging* (pp. 313–343). Erlbaum.

Lachman, R., Shaffer, J. P., & Hennrikus, D. (1974). Language and cognition: Effects of stimulus codability, name-word frequency, and age of acquisition on lexical reaction time. *Journal of Verbal Learning and Verbal Behavior, 13*(6), 613–625. https://doi.org/10.1016/S0022-5371(74)80049-6

La Heij, W., Dirkx, J., & Kramer, P. (1990). Categorical interference and associative priming in picture naming. *British Journal of Psychology, 81*(4), 511–525. https://doi.org/10.1111/j.2044-8295.1990.tb02376.x

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2-3), 259–284. https://doi.org/10.1080/01638539809545028

Lany, J., & Saffran, J. R. (2010). From statistics to meaning: Infants' acquisition of lexical categories. *Psychological Science, 21*(2), 284–291. https://doi.org/10.1177/0956797609358570

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event- related potential study of lexical relationships and prediction in context. *Journal of Memory and Language, 61*(3), 326–338. https://doi.org/10.1016/j.jml.2009.06.004

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience, 25*(3), 484–502. https://doi.org/10.1162/jocn_a_00328

Lau, E. F., Weber, K., Gramfort, A., Hämäläinen, M. S. & Kuperberg, G. R. (2016). Spatiotemporal signatures of lexical–semantic prediction. *Cerebral Cortex, 26*(4), 1377–1387. https://doi.org/10.1093/cercor/bhu219

Lee, E.-K., Brown-Schmidt, S., & Watson, D. G. (2013). Ways of looking ahead: Hierarchical planning in language production. *Cognition, 129*(3), 544–562. https://doi.org/10.1016/j.cognition.2013.08.007

Lee, J. J. & Pinker, S. (2010). Rationales for indirect speech: The theory of the strategic speaker. *Psychological Review, 117*(3), 785–807. https://doi.org/10.1037/a0019688

Lelonkiewicz, J. R., Rabagliati, H., & Pickering, M. J. (2021). The role of language production in making predictions during comprehension. *The Quarterly Journal of Experimental Psychology, 74*(12), 2193–2209. https://doi.org/10.1177/17470218211028438

Lenth, R. (2019). *emmeans: Estimated Marginal Means, aka Least-squares Means.* R package version 1.3.4. https://CRAN.R-project.org/package=emmeans

Levari, T., & Snedeker, J. (2024). Understanding words in context: A naturalistic EEG study of children's lexical processing. *Journal of Memory and Language, 137*, Article 104512. https://doi.org/10.1016/j.jml.2024.104512

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation.* The MIT Press.

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences, 98*(23), 13464–13471. https://doi.org/10.1073/pnas.231459498

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–75. https://doi.org/10.1017/S0140525X99001776

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98*(1), 122–142. https://doi.org/10.1037/0033-295X.98.1.122

Lewis, S. & Phillips, C. (2015). Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research, 44*, 27–46. https://doi.org/10.1007/s10936-014-9329-z

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science, 18*(3), 193–198. https://doi.org/10.1111/j.1467-9280.2007.01871.x

Li, X., Li, X., & Qu, Q. (2022). Predicting phonology in language comprehension: Evidence from the visual world eye-tracking task in Mandarin Chinese. *Journal of Experimental*

Psychology: Human Perception and Performance, 48(5), 531–547. https://doi.org/10.1037/xhp0000999

Li, X., & Qu, Q. (2024). Verbal working memory capacity modulates semantic and phonological prediction in spoken comprehension. *Psychonomic Bulletin & Review, 31*(1), 249–258. https://doi.org/10.3758/s13423-023-02348-5

Liu, Y., Shu, H., & Wei, J. (2006). Spoken word recognition in context: Evidence from Chinese ERP analyses. *Brain and Language, 96*(1), 37–48. https://doi.org/10.1016/j.bandl.2005.08.007

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization.* Oxford University Press.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology, 54*(1), 146–157. https://doi.org/10.1111/psyp.12639

Lukyanenko, C., & Fisher, C. (2016). Where are the cookies? Two- and three-year-olds use number-marked verbs to anticipate upcoming nouns. *Cognition, 146*, 349–370. https://doi.org/10.1016/j.cognition.2015.10.012

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*(1)*,* 19–44. https://doi.org/10.1080/00437956.1959.11659682

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3rd ed.). Lawrence Erlbaum Associates.

Mädebach, A., Jescheniak, J. D., Oppermann, F., & Schriefers, H. (2011). Ease of processing constrains the activation flow in the conceptual-lexical system during speech panning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 649–660. https://doi.org/10.1037/a0022330

Magnusson, J. S., Tanenhaus, M. K., & Aslin, R. N. (2008). Immediate effects of form-class constraints on spoken word recognition. *Cognition*, *108*(3), 866– 873. https://doi.org/10.1016/j.cognition.2008.06.005

Magnusson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (1999). Spoken word recognition in the visual world paradigm reflects the structure of the entire lexicon. *Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society*, 331–336.

Mahon, B., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture–word interference paradigm. Journal of Experimental Psychology: Learning, Memory, and Cognition, 33(3), 503–535. https://doi.org/10.1037/0278-7393.33.3.503

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake — But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 843–847. https://doi.org/10.1037/a0029284

Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: A study of children developing proficiency in reading. *Journal of Experimental Child Psychology, 126*, 264–279. https://doi.org/10.1016/j.jecp.2014.05.004

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS One, 7*(8), e43230. https://doi.org/10.1371/journal.pone.0043230

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1-2), 71–102. https://doi.org/10.1016/0010-0277(87)90005-9

Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is production: The missing link between language production and comprehension. *Scientific Reports, 8*(1), Article 1079. https://doi.org/10.1038/s41598-018-19499-4

Martin, N., Gagnon, D. A., Schwartz, M. F., Dell, G. S., & Saffran, E. M. (1996). Phonological facilitation of semantic errors in normal and aphasic speakers. *Language and Cognitive Processes, 11*(3), 257–282. https://doi.org/10.1080/016909696387187

Martin, N., Weisberg, R. W., & Saffran, E. M. (1989). Variables influencing the occurrence of naming errors: Implications for models of lexical retrieval. *Journal of Memory and Language, 28*(4), 462–485. https://doi.org/10.1016/0749-596X(89)90022-3

Massidda, D. (2013). *retimes: Reaction Time Analysis.* R package version 0.1-2. https://CRAN.R-project.org/package=retimes

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics, 53*(4), 372–380. https://doi.org/10.3758/BF03206780

Mazuka, R., Jincho, N., & Oishi, H. (2009). Development of executive control and language processing. *Language and Linguistics Compass, 3*(1), 59–89. https://doi.org/10.1111/j.1749-818X.2008.00102.x

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. *Proceedings of the 18th Conference of the International Speech Communication Association*, 498–502.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology, 54*(8), 1472–1491. https://doi.org/10.1037/dev0000542

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, *60*(1), 1–39. https://doi.org/10.1016/j.cogpsych.2009.06.003

Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science, 317*(5834), 82. https://doi.org/10.1126/science.1139940

Melinger, A., & Abdel Rahman, R. (2013). Lexical selection is competitive: Evidence from indirectly activated semantic associates during picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 348–364. https://doi.org/10.1037/a0028941

Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition, 35*(3), 494–503. https://doi.org/10.3758/BF03193289

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(6), 1146–1160. https://doi.org/10.1037/0278-7393.17.6.1146

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition, 66*(2), B25 B33. https://doi.org/10.1016/s0010-0277(98)00009-2

Miller, J. F., & Chapman, R. S. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech & Hearing Research, 24*(2), 154–161. https://doi.org/10.1044/jshr.2402.154

Mills, D. L., Prat, C., Zangl, R., Stager, C. L., Neville, H. J., & Werker, J. F. (2004). Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14-and 20-month olds. *Journal of Cognitive Neuroscience, 16*(8), 1–13. https://doi.org/10.1162/0898929042304697

Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye-movements are modulated by reading ability - Evidence from Indian low and high literates. *Journal of Eye Movement Research, 5*(1), Article 3. https://doi.org/10.16910/jemr.5.1.3

Momma, S. (2021). Filling the gap in gap-filling: Long-distance dependency formation in sentence production. *Cognitive Psychology, 129*, 101411. https://doi.org/10.1016/j.cogpsych.2021.101411

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

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 555–563. https://doi.org/10.1037/0278-7393.28.3.555

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*(2), 233–253. https://doi.org/10.2307/747758

Nation, K., Dawson, N. J., & Hsiao, Y. (2022). Book language and its implications for children's language, literacy, and development. *Current Directions in Psychological Science, 31*(4), 375–380. https://doi.org/10.1177/09637214221103264

Nation, K., Marshall, C. M., & Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements.

Journal of Experimental Child Psychology, *86*(4), 314–329. https://doi.org/10.1016/j.jecp.2003.09.001

Navarette, E., Peressotti, F., Lerose, L., & Miozzo, M. (2017). Activation cascading in sign production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(2), 302–318. https://doi.org/10.1037/xlm0000312

Navarrete, E., & Costa, A. (2009). The distractor picture paradox in speech production: Evidence from the word translation task. *Journal of Psycholinguistic Research, 38*(6), 527–547. https://doi.org/10.1007/s10936-009-9119-1

Ng, S., Payne, B. R., Stine-Morrow, E. A. L., & Federmeier, K. D. (2018). How struggling adult readers use contextual information when comprehending speech: Evidence from event-related potentials. *International Journal of Psychophysiology, 125*(3), 1–9. https://doi.org/10.1016/j.ijpsycho.2018.01.013

Nilsen, E. S. & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology, 58*(2), 220–249. https://doi.org/10.1016/j.cogpsych.2008.07.002

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*(3), 189–234. https://doi.org/10.1016/0010-0277(94)90043-4

Nyström, M., Niehorster, D. C., Andersson, R., & Hooge, I. (2021). The Tobii Pro Spectrum: A useful tool for studying microsaccades? *Behavior Research Methods, 53*(1), 335–353. https://doi.org/10.3758/s13428-020-01430-3

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 17*(4), 273–281. https://doi.org/10.1080/17470216508416445

Ovans, Z. (2022). *Developmental parsing and cognitive control* [Doctoral dissertation, University of Maryland, College Park]. Digital Repository at the University of Maryland. https://doi.org/10.13016/en2r-ce6z

Özge, D., Kornfilt, J., Maquate, K., Küntay, A. C., & Snedeker, J. (2022). German-speaking children use sentence-initial case marking for predictive language processing at age four. *Cognition, 221*, Article 104988. https://doi.org/10.1016/j.cognition.2021.104988

Özge, D., Küntay, A., & Snedeker, J. (2019). Why wait for the verb? Turkish speaking children use case markers for incremental language comprehension. *Cognition*, *183*, 152–180. https://doi.org/10.1016/j.cognition.2018.10.026

Paivio, A., Clark, J. M., Digdon, N., & Bons, T. (1989). Referential processing: Reciprocity and correlates of naming and imaging. *Memory & Cognition, 17*(2), 163–174. https://doi.org/10.3758/BF03197066

Panichello, M. F., Cheung, O.S., & Bar, M. (2013). Predictive feedback and conscious visual experience. *Frontiers in Psychology, 3,* Article 620. https://doi.org/10.3389/fpsyg.2012.00620

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable webcam eye tracking using user interactions. *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI), 25*(1), 3839–3845.

Passler, M. A., Isaac, W., & Hynd, G. W. (1985). Neuropsychological development of behavior attributed to frontal lobe functioning in children. *Developmental Neuropsychology, 1*(4), 349–370. https://doi.org/10.1080/87565648509540320

Pattamadilok, C., Knierim, I. N., Duncan, K. J. K., & Devlin, J. T. (2010). How does learning to read affect speech perception? *The Journal of Neuroscience, 30*(25), 8435–8444. https://doi.org/10.1523/JNEUROSCI.5791-09.2010

Paul, P., Ziegler, J., Chalmers, E., & Snedeker, J. (2019). Children and adults successfully comprehend subject-only sentences online. *PLoS ONE, 14*(1), Article e0209670. https://doi.org/10.1371/journal.pone.0209670

Perre, L., Pattamadilok, C., Montant, M., & Ziegler, J. C. (2009). Orthographic effects in spoken language: On-line activation or phonological restructuring? *Brain Research, 1275*, 73–80. https://doi.org/10.1016/j.brainres.2009.04.018

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(3), 539–557. https:// doi.org/10.1037/0278-7393.24.3.539

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Pickering, M. J. & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Science, 11*(3), 105–110. https://doi.org/10.1016/j.tics.2006.12.002

Pickering M. J., Garrod S. (2013a). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4), 329–347. https://doi.org/10.1017/s0140525x12001495

Pickering, M. J., & Garrod, S. (2013b). How tightly are production and comprehension interwoven? *Frontiers in Psychology, 4*, Article 238. https://doi.org/10.3389/fpsyg.2013.00238

Poarch, G. J., & van Hell, J. G. (2012). Cross-language activation in children's speech production: Evidence from second language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology, 111*(3), 419–438. https://doi.org/10.1016/j.jecp.2011.09.008

Prystauka, Y., Altmann, G. T., & Rothman, J. (2023). Online eye tracking and real-time sentence processing: On opportunities and efficacy for capturing psycholinguistic effects of different magnitudes and diversity. *Behavioral Research Methods, 56*, 3504–3522. https://doi.org/10.3758/s13428-023-02176-4

Rabagliati, H., Gambi, C., & Pickering, M. J. (2016). Learning to predict or predicting to learn? *Language, Cognition and Neuroscience, 31*(1), 94–105. https://doi.org/10.1080/23273798.2015.1077979

Rabagliati, H., Pylkkänen, L., & Marcus, G. F. (2013). Top-down influence in young children's linguistic ambiguity resolution. *Developmental psychology*, *49*(6), 1076–1089. https://doi.org/10.1037/a0026918

Rabs, E., Delogu, F., Drenhaus, H., & Crocker, M. W. (2022). Situational expectancy or association? The influence of event knowledge on the N400. *Language, Cognition and Neuroscience*, *37*(6), 766–784. https://doi.org/10.1080/23273798.2021.2022171

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*(3), 460– 499. https://doi.org/10.1037/0033-295X.107.3.460

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86*(3), 446–461. https://doi.org/10.1037/0033-2909.86.3.446

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*(3), 510–532. https://doi.org/10.1037/0033-2909.114.3.510

Rayner, K., Slowiaczek, M. L., Clifton, C., & Bertera, J. H. (1983). Latency of sequential eye movements: Implications for reading. *Journal of Experimental Psychology: Human Perception and Performance, 9*(6), 912–922. https://doi.org/10.1037/0096-1523.9.6.912

R Core Team. (2021). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reuter, T., Borovsky, A., & Lew-Williams, C. (2019). Predict and redirect: Prediction errors support children's word learning. *Developmental Psychology, 55*(8), 1656–1665. https://doi.org/10.1037/dev0000754

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research, 53*(2), 333–349. https://doi.org/10.1044/1092-4388(2009/08-0183)

Ridderinkhof, K. R. (2002). Activation and suppression in conflict tasks: Empirical clarification through distributional analyses. In W. Prinz & B. Hommel (Eds.), *Attention and Performance XIX: Common Mechanisms in Perception and Action* (pp. 494–519). Oxford University Press.

Riggs, K. J., Mctaggart, J., Simpson, A., & Freeman, R. P. (2006). Changes in the capacity of visual working memory in 5- to 10-year-olds. *Journal of Experimental Child Psychology, 95*(1), 18–26. https://doi.org/10.1016/j.jecp.2006.03.009

Rigler, H., Farris-Trimble, A., Greiner, L., Walker, J., Tomblin, J. B., & McMurray, B. (2015). The slow developmental time course of real-time spoken word recognition. *Developmental Psychology, 51*(12), 1690–1703. https://doi.org/10.1037/dev0000044

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*(1-3), 107–142. https://doi.org/10.1016/0010-0277(92)90041-F

Roelofs, A. (2008). Tracing attention and the activation flow of spoken word planning using eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(2), 353–368. https://doi.org/10.1037/0278-7393.34.2.353

Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1996). Interaction between semantic and orthographic factors in conceptually driven naming: Comment of Starreveld and LaHeij (1995). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), 246–251. https://doi.org/10.1037/0278-7393.22.1.246

Rossion, B., Caldara, R., Seghier, M., Schuller, A.-M., Lazeyras, F., & Mayer, E. (2003). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain, 126*(11), 2381–2395. https://doi.org/10.1093/brain/awg241

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33*(2), 217–236. https://doi.org/10.1068/p5117

Ryskin, R., & Nieuwland, M. S. (2023). Prediction during language comprehension: What is next? *Trends in Cognitive Sciences, 27*(11), 1032–1052. https://doi.org/10.1016/j.tics.2023.08.003

Sachs, J. & Devin, J. (1976). Young children's use of age-appropriate speech styles in social interaction and role-playing. *Journal of Child Language, 3*(1), 81–98. https://doi.org/10.1017/S030500090000132X

Sadat, J., Martin, C. D., Costa, A., & Alario, F.-X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology, 68*, 33–58. https://doi.org/10.1016/j.cogpsych.2013.10.001

Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America, 57*(8), 1030–1033. https://doi.org/10.1364/JOSA.57.001030

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*(6), e13335. https://doi.org/10.1111/psyp.13335

Schipley, K.G., & McAfee, J. G. (2015). *Assessment in speech-language pathology: A resource manual* (5th edition). Cengage Learning.

Schneider, W., & Bjorklund, D. F. (1998). Memory. In W. Damon, D. Kuhn, & R. S. Siegler (Eds.), *Handbook of Child Psychology, Vol. 2: Cognitive, Language, and Perceptual Development* (5th ed., pp. 467–521). Wiley.

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*(1), 86–102. https://doi.org/10.1016/0749-596X(90)90011-N

Schriefers, H., & Vigliocco, G. (2015). Speech production, psychology of. In J. D. Wright (Ed.), *International Encyclopedia of the Social and Behavioral Sciences* (2nd ed.). (pp. 255–258). Elsevier.

Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two step model of lexical access: Evidence from picture naming. *Journal of Memory and Language, 54*(2), 228–264. https://doi.org/10.1016/j.jml.2005.10.001

Seidl, A., Hollich, G., & Jusczyk, P. W. (2003). Early Understanding of Subject and Object Wh-Questions. *Infancy*, *4*(3), 423–436. https://doi.org/10.1207/S15327078IN0403_06

Sekerina, I. A., & Brooks, P. J. (2007). Eye movements during spoken word recognition in Russian children. *Journal of Experimental Child Psychology, 98*(1), 20–45. https://doi.org/10.1016/j.jecp.2007.04.005

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavioral Research Methods, 50*(2), 451–465. https://doi.org/10.3758/s13428-017-0913-7

Shipley, K. G., & McAfee, J. G. (2015). *Assessment in Speech Language Pathology: A Resource Manual* (5th ed.). Cengage Learning.

Sieger-Gardner, L., & Schwartz, R. G. (2008). Lexical access in children with and without specific language impairment: A cross–model picture–word interference study. *International Journal of Language & Communication Disorders, 43*(5), 528–551. https://doi.org/10.1080/13682820701768581

Silva-Pereyra, J. F., Klarman, L., Lin, L. J-F., & Kuhl, P. K. (2005). Sentence processing in 30-month-old children: An event-related potential study. *NeuroReport, 16*(6), 645–648. https://doi.org/10.1097/00001756-200504250-00026

Silva-Pereyra, J., Rivera-Gaxiola, M., & Kuhl, P. K. (2005). An event-related brain potential study of sentence comprehension in preschoolers: Semantic and morphosyntactic processing. *Cognitive Brain Research, 23*(2–3), 247–258. https://doi.org/10.1016/j.cogbrainres.2004.10.015

Simmering, V. R. (2012). The development of visual working memory capacity during early childhood. *Journal of Experimental Child Psychology, 111*(4), 695–707. https://doi.org/10.1016/j.jecp.2011.10.007

Simmons, E. S. (2017). The timecourse of phonological competition in spoken word recognition: A comparison of adults and very young children [Master's thesis, University of Connecticut]. Retrieved from https://opencommons.uconn.edu/gs_theses/1156/.

Slim, M. S., & Hartsuiker, R. J. (2022). Moving visual world experiments online? A web-based replication of Dijkgraaf, Hartsuiker, and Duyck (2017) using PCIbex and WebGazer.js. *Behavioral Research Methods, 55*, 3786–3804. https://doi.org/10.3758/s13428-022-01989-z

Slim, M. S., Hartsuiker, R. J., & Snedeker, J. (2022). The real-time resolution of quantifier scope ambiguity. Paper presented at the 22nd ESCOP Conference, Université de Lille, Lille, France.

Slim, M. S., Kandel, M., Yacovone, A., & Snedeker, J. (2024). Webcams as windows to the mind? A direct comparison between in-lab and web-based eye-tracking methods. *Open Mind: Discoveries in Cognitive Science, 8*, 1369–1424. https://doi.org/10.1162/opmi_a_00171

Snedeker, J. (2009). Children's Sentence Processing. In E. Bavin (Ed.), *The Handbook of Child Language* (pp. 331–338). Cambridge University Press. http://dx.doi.org/10.1017/S0305000910000115

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299. https://doi.org/10.1016/j.cogpsych.2004.03.001

Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language*, *58*(2), 574–608. https://doi.org/10.1016/j.jml.2007.08.001

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Sommerfeld, L., Staudte, M., Mani, N., & Kray, J. (2023). Even young children make multiple predictions in the complex visual world. *Journal of Experimental Child Psychology*, *235*, Article 105690. https://doi.org/10.1016/j.jecp.2023.105690

Sorger, B., Goebel, R., Schiltz, C., & Rossion, B. (2007). Understanding the functional neuroanatomy of acquired prosopagnosia. *NeuroImage, 35*(2), 836–852. https://doi.org/10.1016/j.neuroimage.2006.09.051

Spieler, D. H., & Griffin, Z. M. (2006). The influence of age on the time course of word preparation in multiword utterances. *Language and Cognitive Processes, 21*(1-3), 291–321. https://doi.org/10.1080/01690960400002133

SR Research. (2021). *EyeLink® 1000 Plus Brochure.*

Stafford, T., & Gurney, K. N. (2011). Additive factors do not imply discrete processing stages: A worked example using models of the Stroop task. *Frontiers in Psychology, 2*, Article 287. https://doi.org/10.3389/fpsyg.2011.00287

Starreveld, P. A. (2000). On the interpretation of onsets of auditory context effects in word production. *Journal of Memory and Language, 42*(4), 497–525. https://doi.org/10.1006/jmla.1999.2693

Starreveld, P. A., & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(3), 686–698. https://doi.org/10.1037/0278-7393.21.3.686

Staub, A. (2010). Response time distributional evidence for distinct varieties of number attraction. *Cognition, 114*(3), 447–454. https://doi.org/10.1016/j.cognition.2009.11.003

Stemberger, J. P. (1983). Inflectional malapropisms: Form based errors in English morphology. *Linguistics, 21*, 573–602. https://doi.org/10.1515/ling.1983.21.4.573

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. Acta *Psychologica, 30*, 276–315. https://doi.org/10.1016/0001-6918(69)90055-9

Sternberg, S. (1984). Stage models of mental processing and the additive-factor method. *Behavioral and Brain Sciences, 7*(1), 55–94. https://doi.org/10.1017/S0140525X00026285

Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *An Invitation to Cognitive Science: Methods, Models, and Conceptual Issues* (Vol. 4) (pp. 703–863). MIT Press.

Sternberg, S. (2001). Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychologica, 106*(1-2), 147–246. https://doi.org/10.1016/S0001-6918(00)00045-7

Strand, J. F., Brown, V. A., Brown, H. E., & Berg, J. J. (2018). Keep listening: Grammatical context reduces but does not eliminate activation of unexpected words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(6), 962–973. https://doi.org/10.1037/xlm0000488

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: Electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex, 20*(4), 912–928. https://doi.org/10.1093/cercor/bhp153

Swingley, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language*, *60*(2), 252–269. https://doi.org/10.1016/j.jml.2008.11.003

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*(2), 73-108. https://doi.org/10.1016/s0010-0277(99)00021-9

Sylvia, L. (2017). *Lexical access in children with specific language impairment and deficits in auditory processing.* [Doctoral dissertation, City University of New York]. CUNY Academic Works.

Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., Arévalo, A., & Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers, 35*(4), 621–633. https://doi.org/10.3758/BF03195542

Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., Arévalo, A., Vargha, A., & Bates, E. (2005). Timed action and object naming. *Cortex, 41*(1), 7–25. https://doi.org/10.1016/S0010-9452(08)70174-6

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

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science, 9*(5), 379–385. https://doi.org/10.1111/1467-9280.00071

Thomas, R. D. (2006). Processing time predictions of current models of perception in the classic additive factors paradigm. *Journal of Mathematical Psychology, 50*(5), 441–455. https://doi.org/10.1016/j.jmp.2006.05.006

Thothathiri, M., & Snedeker, J. (2008). Syntactic priming during language comprehension in three- and four-year-old children. *Journal of Memory and Language, 58*(2), 188–213. https://doi.org/10.1016/j.jml.2007.06.012

Tiffin-Richards, S. P., & Schroeder, S. (2020). Context facilitation in text reading: A study of children's eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(9), 1701–1713. https://doi.org/10.1037/xlm0000834

Tincoff, R., & Juscyzk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10*(2), 172–175. https://doi.org/10.1111/1467-9280.00127

Tincoff, R., & Juscyzk, P. W. (2011). Six-month-olds comprehend words that refer to parts of the body. *Infancy, 17*(4), 432–444. https://doi.org/10.1111/j.1532-7078.2011.00084.x

Tobii (2010). *Tobii TX300 Eye Tracker.*

Tobii Pro (2021). *Pro Spectrum User Manual.*

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*, 89–134. https://doi.org/10.1016/s0010-0277(99)00032-3

Valenti, R., Staiano, J., Sebe, N., & Gevers, T. (2009). Webcam-based visual gaze estimation. *International Conference on Image Analysis and Processing– ICIAP 2009, 5716,* 662–671. https://doi.org/10.1007/978-3-642-04146-4_71

Valliappan, N., Dai, N., Steinberg, E., He, J., Rogers, K., Ramachandran, V., Xu, P., Shojaeizadeh, M., Guo, L., Kohlhoff, K., & Navalpakkam, V. (2020). Accelerating eye movement research via accurate and affordable smartphone eye tracking. *Nature Communications, 11*(1), 4553. https://doi.org/10.1038/s41467-020-18360-5

Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes, 8*(4), 485–531. https://doi.org/10.1080/01690969308407586

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology, 48*(4), 422–488. https://doi.org/10.1016/j.cogpsych.2003.09.001

Vincent, S. B. (1912). The function of the viborissae in the behavior of the white rat. *Behavioral Monographs, 1*(5), 81.

Vos, M., Minor, S., & Ramchand, G. C. (2022). Comparing infrared and webcam eye tracking in the Visual World Paradigm. *Glossa Psycholinguistics, 1*(1), https://doi.org/10.5070/G6011131

Vurpillot, E., & Ball, W. A. (1979). The concept of identity and children's selective attention. In G. A. Hale & M. Lewis (Eds.), *Attention and Cognitive Development* (pp. 23–42). Plenum.

Wagner V., Jescheniak J. D., Schriefers H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 423–440. https://doi.org/10.1037/a0018619

Waite, B., Yacovone, A., & Snedeker, J. (under review). Case clozed: Young children can explicitly predict upcoming words in a naturalistic, story-based cloze task.

Walker, R., Walker, D. G., Husain, M., & Kennard, C. (2000). Control of voluntary and reflexive saccades. *Experimental Brain Research, 130*(4), 540–544. https://doi.org/10.1007/s002219900285

Wang, L., Brothers, T., Jensen, O., & Kuperberg, G. (2024). Dissociating the pre-activation of word meaning and form during sentence comprehension: Evidence from EEG representational similarity analysis. *Psychonomic Bulletin & Review, 31,* 862–873. https://doi.org/10.3758/s13423-023-02385-0

Weighall, A. R., Henderson, L. M., Barr, D. J., Cairney, S. A., Gaskell, M. G. (2017). Eye-tracking the time-course of novel word learning and lexical competition in adults and children. *Brain and Language*, *167*, 13–27. https://doi.org/10.1016/j.bandl.2016.07.010

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology, 7*(2), 131–149. https://doi.org/10.1080/87565649109540483

Werker, J. F., Fennel, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy, 3*(1), 1–30. https://doi.org/10.1207/S15327078IN0301_1

White, C. T., Eason, R. G., & Bartlett, N. R. (1962). Latency and duration of eye movements in the horizontal plane. *Journal of the Optical Society of America, 52*(2), 210–213. https://doi.org/10.1364/josa.52.000210

Wiebe S. A., Sheffield T. D., & Espy K. A. (2012). Separating the fish from the sharks: A longitudinal study of preschool response inhibition. *Child Development*, *83*(4), 1245–1261. https://doi.org/10.1111/j.1467-8624.2012.01765.x

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage, 62*(1), 356–366. https://doi.org/10.1016/j.neuroimage.2012.04.054

Wolfe, M. B. W., & Goldman, S. R. (2003). Use of latent semantic analysis for predicting psychological phenomena: Two issues and proposed solutions. *Behavior Research Methods, Instruments & Computers, 35*(1), 22–31. https://doi.org/10.3758/BF03195494

Wyatte, D., Jilk, D. J., & O'Reilly, R. C. (2014). Early recurrent feedback facilitates visual object recognition under challenging conditions. *Frontiers in Psychology, 5,* Article 674. https://doi.org/10.3389/fpsyg.2014.00674

Xu, P., Ehinger, K. A., Zhang, Y., Finkelstein, A., Kulkarni, S. R., & Xiao, J. (2015). TurkerGaze: Crowdsourcing saliency with webcam based eye tracking. *arXiv*. http://arxiv.org/abs/1504.06755

Yacovone, A., Shafto, C. K., Worek, A., & Snedeker, J. (2021). World vs. world knowledge: A developmental shift from bottom-up lexical cues to top-down plausibility. *Cognitive Psychology, 131*, Article 101442. https://doi.org/10.1016/j.cogpsych.2021.101442

Yacovone, A., Waite, B., Levari, T., & Snedeker, J. (2024). Let them eat ceke: An electrophysiological study of form-based prediction in rich naturalistic contexts. *Journal of Experimental Psychology: General.* Advance online publication. https://doi.org/10.1037/xge0001677

Yang, Q., Bucci, M. P., & Kapoula, Z. (2002). The latency of saccades, vergence, and combined eye movements in children and in adults. *Investigative Ophthalmology & Visual Science, 43*(9), 2939–2949.

Yang, X., & Krajbich, I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making, 16*(6), 1485–1505. https://doi.org/10.1017/S1930297500008512

Yang, T.-X., Xie, W., Chen, C.-S., Altgassen, M., Wang, Y., Cheung, E. F. C., & Chan, R. C. K. (2017). The development of multitasking in children ages 7–12 years: Evidence from cross-sectional and longitudinal data. *Journal of Experimental Child Psychology, 161*, 63–80. https://doi.org/10.1016/j.jecp.2017.04.003

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(1), 1–14. https://doi.org/10.1037/0278-7393.32.1.1

Zangl, R., Klarman, L., Thal, D., Fernald, A., & Bates, E. (2005). Dynamics of word comprehension in infancy: Developments in timing, accuracy, and resistance to acoustic degradation. *Journal of Cognition and Development, 6*(2), 179–208. https://doi.org/10.1207/s15327647jcd0602_2

Zareva, A., Schwanenflugel, P. & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: variable sensitivity. *Studies in Second Language Acquisition*, 27(4), 567–595. https://doi.org/10.1017/S0272263105050254

Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX).* https://doi.org/10.17605/OSF.IO/MD832

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word Reading and other tasks. *Journal of Memory and Language, 47*(1), 1–29. https://doi.org/10.1006/jmla.2001.2834

Zhang, H., Carlson, M. T., & Diaz, M. T. (2020). Investigating effects of phonological neighbors on word retrieval and phonetic variation in word naming and picture naming paradigms. *Language, Cognition, and Neuroscience, 35*(8), 980–991. https://doi.org/10.1080/23273798.2019.1686529

Zhang, Q., Zhu, X., & Damian, M. (2018). Phonological activation of category coordinates in spoken word production: Evidence for cascaded processing in English but not in

Mandarin. *Applied Psycholinguistics, 39*(5), 835–860. https://doi.org/10.1017/S0142716418000024

Zhou, P., Crain, S., & Zahn, L. (2014). Grammatical aspect and event recognition in children's online sentence comprehension. *Cognition, 133*(1), 262–276. https://doi.org/10.1016/j.cognition.2014.06.018.