UNIVERSITY OF CALIFORNIA

Los Angeles

Extralinguistic Cognition and Verb Argument Structure in Development: Learning and Processing

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Linguistics

by

Ekaterina Andreevna Khlystova

2024

ABSTRACT OF THE DISSERTATION

Extralinguistic Cognition and Verb Argument Structure in Development: Learning and Processing

by

Ekaterina Andreevna Khlystova
Doctor of Philosophy in Linguistics
University of California, Los Angeles, 2024
Professor Laurel L. Perkins, Chair

This dissertation investigates the interaction of developing extralinguistic cognitive systems with early language learning and processing through the case study of verb argument structure. The interaction of these systems with the linguistic system underpins fundamental theories of language learning and use: language does not exist in isolation. The extralinguistic cognitive systems supporting language learning and use are themselves in their earliest stages of development, potentially placing perceptual and computational constraints on learning. At later stages of development, when children have already learned a significant amount about their target language(s), the use of this acquired knowledge can further be influenced and constrained by still-developing cognitive systems.

We approach this complex interaction through two case studies examining the role of extralinguistic cognitive systems in verb argument structure (i) acquisition and (ii) processing. First, we examine how perceptual systems like visual working memory, which are limited throughout development, can support the representation of high-adicity event concepts. In two behavioral studies, we investigate whether the visual working memory system of adults and preschool-aged children is capable of yielding a 4-participant event concept. We show that adults are capable of representing a trading scene under a single, 4-place concept, despite the typical limitations of the visual working memory system. We also show that preschool-aged children are capable of representing all 4 characters in the same trading scene as participants, raising questions of how young learners may map between their conceptual and linguistic representations. Second, we computationally investigate how immature extralinguistic cognitive systems, such as working memory and cognitive control, can impact the deployment of verb argument structure knowledge. To do so, we develop and implement a generalized left-corner parser with independent parameters corresponding to these systems. We successfully model parsing performance in a well-known example where children's parsing differs from that of adults', generating testable empirical predictions that could further adjudicate between these two systems.

By investigating independently developing cognitive systems through the lens of verb argument structure, this dissertation explores two ways that these systems can influence verb learning and use. In doing so, we provide an example of how to isolate the specific contributions of extralinguistic cognition to the acquisition and deployment of one type of linguistic knowledge, which can generalize beyond the realm of verb argument structure to other aspects of language development and acquisition.

The dissertation of Ekaterina Andreevna Khlystova is approved.

Timothy Hunter

Jeffrey L. Lidz

Megha Sundara

Laurel L. Perkins, Committee Chair

University of California, Los Angeles

2024

To my mother, who sacrificed her own
doctorate to give us a head start in life.
Thank you for everything.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Words cannot begin to express the amount of gratitude I have for each person who helped make this dissertation possible. In keeping with tradition, however, I must present an abridged version of my gratitude for everybody who has impacted my journey towards this Ph.D.:

I have been unbelievably lucky to have had some incredible mentors and advisors over the course of my graduate career. Thank you, Megha, for taking me under your wing as a bright-eyed, bushy-tailed first year student, coming straight from a Chemistry degree and having minimal knowledge of the world of theoretical and experimental Linguistics. I am grateful for all of the Skype meetings we had that first year while you were on sabbatical, before remote work became the norm in 2020. Thank you for being a calming and inspiring presence when my entire Master's plan was derailed by a global pandemic, and introducing me to the corpus project which has truly been foundational to the researcher I have become today. Thank you for always being willing to take time to chat, even when I decided to pursue a completely different direction for this dissertation, and as you became ever-busier with all-new responsibilities as department chair. And thank you for always taking the time to celebrate the wins, no matter how big or small.

I also owe an incredible debt of gratitude to my advisor, Laurel. Thank you for encouraging me to develop and pursue a seemingly tangential interest (at the time) in child sentence processing during our very first meeting in early 2020. Thank you so much for all of the support and encouragement you've provided as I've tackled this dissertation, for pushing me to think deeply and have the courage to attack some of the most mind-bending puzzles in language acquisition, and guiding me as I stumbled on various speed bumps and obstacles along the way. Thank you, as well, for the many giggles we shared as we discussed the challenges of working with infants and preschoolers – I don't think I could have become such a "preschooler-whisperer" if I hadn't been coached by the master herself! And thank you for your patience, even when I didn't know much of anything. It has truly been an honor being your first doctoral student, and I look forward to joining an incredible legacy of future academic siblings. I am confident your brilliance, combined with your excellence in mentorship, will foster some amazing new minds in language acquisition research, and I look forward to seeing the plethora of brilliant researchers you will advise.

I have also had the honor of having some truly brilliant committee members. Thank you to Tim, for pushing me to think strategically and deeply about parsing. Thank you as well, for the many hours you spent helping me debug my code when I was near tears of desperation trying to find the source of my errors. And thank you for spending so much time perfecting the lookahead algorithm. Thank you to Jeff, for our discussions as I took on the challenge of learning about event perception and verb learning. And thank you for sharing your wisdom on all things language acquisition – your encyclopedic knowledge of the field is truly awe-inspiring.

I've been incredibly fortunate to have made some lifelong friends along the way. I'm grateful to Iza, Lily, and Kiki, for experiencing the trials and tribulations of infant and toddler testing alongside me these past few years – there's never a boring day, and I'm glad we were able to share laughs as well as troubleshoot frustrations throughout the whole experience. Thank you to Huilei for your cheerful smile and calming presence – you never failed to brighten my day. Yang – thank you for becoming a lifelong friend through this experience. They say the people you dissertate with are people you stick with, and I am incredibly grateful to have had you on the same timeline as me as we scrambled to tie our dissertations together by the end of our fifth years. I'm so incredibly proud of you. Finally, thank you to Candace (my favorite cabbage), for being the best friend anyone could ask for. I'm so glad our dogs introduced us 3 years ago.

Thank you to my family, who instilled in me the drive to learn about as many things as I can early on. Thank you for always pursuing the highest degrees of excellence and inspiring me to do so in my own right, and for giving me the confidence to pursue a novel, "soft" science. And thank you for tolerating my endless stories about the infants and preschoolers I've worked with over the course of graduate school.

Lastly, thank you to Evan; I couldn't have done this without you. Here's to forever taking on challenges together.

VITA

2023-2025	NSF Doctoral Dissertation Research Improvement (DDRI) grant
2024	Grad Slam Finalist, UCLA
2023-2024	Dissertation Year Fellowship
2022	Keck Summer Research Fellowship
2021-2022	Graduate Research Mentorship Fellow
2021	M.A. (Linguistics), UCLA, Los Angeles, C.A.
2021	Graduate Summer Research Mentorship Fellow
2019	B.S. (Chemistry) and B.A. (Linguistics), Duke University, Durham, N.C.
2015-2019	Angier B. Duke Memorial Scholarship, Duke University, Durham, N.C.

PUBLICATIONS AND PRESENTATIONS

Khlystova, E., Williams, A., Lidz, J., and Perkins, L. (2024). "Visual perception supports 4-place event representations: A case study of TRADING." Proceedings of the Annual Meeting of the Cognitive Science Society.

Khlystova, E., Chong, A. J., and Sundara, M. (2023). "Phonetic variation in English infant-directed speech: A large-scale corpus analysis". Journal of Phonetics, 100, 101267.

Khlystova, E., Williams, A., Lidz, J., and Perkins, L. (2024). “Visual perception supports 4-place event representations: A case study of TRADING.” Annual Meeting of the Cognitive Science Society: Talk, July 2024. Rotterdam, Netherlands.

Khlystova, E. and Sundara, M.. “The lexical representation of coda /t/”. Acoustical Society of America (ASA): Poster presentation, Dec 2023. Sydney, Australia.

Narkar, J., Khlystova, E., Mayer, C., Aly, A., Kim, J. Y., and Sundara, M.. “Evaluating the learnability of vowel categories from Infant-Directed Speech”. Acoustical Society of America (ASA): Poster presentation, May 2022. Denver, CO.

Khlystova, E., Chong, A., and Sundara, M.. “Quantifying phonetic variation: A large-scale corpus analysis of coronal segments in English infant-directed speech”. Acoustical Society of America (ASA): Poster presentation, Nov 2021. Seattle, WA.

Khlystova, E., Chong, A. J., and Sundara, M.. “Phonetic variation in coronals in English infant-directed speech: A large-scale corpus analysis”. Boston University Conference on Language Development (BUCLD): Poster presentation, Nov 2021. Virtual.

CHAPTER 1

Introduction

Language acquisition, like many processes in development, occurs in tandem with the development of other cognitive capacities. As a result, the process of language learning is shaped by a dynamic and developing cognitive landscape. How might the development of these extralinguistic cognitive systems influence early language learning?

In and of itself, language acquisition poses a fundamental logical problem: in order to fully acquire a language, the young learner must generalize beyond a finite set of data to a system that can generate an infinite number of linguistic expressions (Chomsky, 1965). Traditional theories of language acquisition posit that children draw on a finite dataset (their linguistic input), then use their cognitive abilities and an initial hypothesis space to draw generalizations about the properties of the language(s) they are learning (Chomsky, 1965; Wexler & Culicover, 1980). For this reason, language acquisition research has typically approached language learning from two perspectives, probing both the initial hypothesis space from which the learner approaches language acquisition, and the statistical regularities in the language input which may provide evidence for particular hypotheses over others within that hypothesis space.

Infants and children are indeed highly sensitive to the statistical regularities in their input (for a review, see Lidz & Gagliardi, 2015). But traditional theories of language acquisition have typically abstracted away from the impact that still-ongoing development of extralinguistic cognitive systems can have on language learning (Chomsky, 1965; Wexler & Culicover, 1980). Similarly, these theories have also abstracted away from the fact that children's representations of the language input they receive also necessarily changes throughout development as they acquire more linguistic knowledge. Crucially, language learning is an iterative process, building upon imperfect, existing knowledge at each developmental time point to adjudicate within the hypothesis space and make further generalizations (Lidz & Gagliardi, 2015; Perkins et al., 2022). For example, children who don't yet know any words in their language may struggle to learn that their language is a subject-drop language. At early stages of language learning, the patterns that may appear obvious to adult language users are liable to be obscured by a lack of other relevant linguistic knowledge. Consequently, the language input a child hears may not be representative of the *intake* from which the child learns (Fodor, 1998; Lidz & Gagliardi, 2015; Perkins et al., 2022; Valian, 1990, among others).

Moreover, children's language learning is supported by cognitive systems that are developing in and of themselves, potentially constraining or obscuring important regularities in the language input from which generalizations can be drawn. To learn from their language input, young learners must be able to process the sentences being spoken around them and extract relevant linguistic information. Similarly, the young learner must also be capable of sufficiently representing the nonlinguistic contexts in which language is used: these contexts provide a space of possibilities for the meanings of words and the sentences that learners hear. Importantly, learning from linguistic and conceptual representations requires support from extralinguistic cognitive systems, such as working memory and inhibitory control. These cognitive systems are, however, in the process of developing themselves, which can place constraints on the types of representations that serve as the foundation of acquiring meaning and structure in language. For example, understanding sentences in real time requires rapid, incremental processing as each word in the sentence unfolds. The presence of ambiguity in language (including structural ambiguity) means real-time understanding requires the learner to maintain multiple possible interpretations simultaneously. But if immature cognitive systems prevent the consideration of less probable analyses, the language input runs the risk of not being parsed veridically, leading to additional complications in language learning (Lidz, 2023; Lidz et al., 2017; Omaki & Lidz, 2015; White & Lidz, 2022).

A significant body of work has investigated the ways two developing extralinguistic cognitive systems in particular – working memory and cognitive control – can influence various processes over the course of language development (Hsu et al., 2021; Montgomery et al., 2008; Ness et al., 2022; Ovans, 2022; Trueswell et al., 1999; Weighall & Altmann, 2011; Woodard et al., 2016; Zhou et al., 2021, among others). In this dissertation, I investigate the interaction of these two systems with language development through the lens of verb argument structure. In particular, we investigate how the architecture of developing memory systems, in conjunction with other executive functions, can impact 1) the way that young learners represent the events that are labeled by verbs and 2) the way that young learners deploy knowledge of verb argument structure to process sentences in real time.

1.1 Extralinguistic cognition in argument structure learning

One crucial aspect of language development entails identifying verb argument structure (the number and type of arguments, such as subject and object, that a verb requires). For example, language learners must learn whether a given verb is intransitive (can only take a subject; e.g., *run*), transitive (requires a subject and an object; e.g., *kick*), optionally transitive (requires a subject and can optionally take an object; e.g., *eat*), ditransitive (requires a subject and both a direct and indirect object; e.g., *put*), and so on. In order to acquire a verb, young learners must not only learn the verb's argument structure, but also map between their conceptual representation of a given event and their representation of the linguistic structure used to label that event. There are systematic correlations between a verb's distribution with clause arguments and its meaning (Dowty, 1991; Fillmore, 1968, 1972; Jackendoff, 1990), and there have been many proposals for how children may achieve this mapping, including syntactic bootstrapping (Gleitman, 1990; Landau & Gleitman, 1985) and semantic bootstrapping (Grimshaw, 1981; Pinker, 1984, 1989). In syntactic bootstrapping, infants are thought to exploit their existing knowledge of argument structure to identify the relevant participants in a speaker's view of the event that a sentence is describing, in order to learn the meaning of the verb in that sentence. Semantic bootstrapping, on the other hand, purports that infants use their conceptual understanding of a speaker's view of an event to make inferences about syntax (and proponents of these theories have further argued that semantic bootstrapping can extend to verb argument structure learning; see Pinker, 1989). However, regardless of whether infants are learning structure from meaning or meaning from structure, any sort of structure-meaning mapping requires the learner to relate the clause arguments they hear to the conceptual representations under which they view the particular events in the world. But what, exactly, are children's conceptual representations of events?

Bootstrapping theories that posit children learn verbs by mapping between event participants and clause arguments assume that the learner views scenes under concepts in which the relevant event participants are explicitly represented. In language learners with typical vision, viewing an event yields a conceptual structure from which the learner can map participants to arguments (Hafri & Firestone, 2021; Hafri et al., 2013; Hafri et al., 2018). This requires the learner to visually track participants, which necessarily implicates the visual working memory system. However, the visual working memory literature highlights a potential obstacle to learning verbs with complex argument structures, such as those that label 4-place predicates (i.e., verbs like *trade*): across all stages of development, humans have a visual working memory limit of 3-4 items (Cowan, 2001; Halberda et al., 2006; Scholl & Pylyshyn, 1999; Sperling, 1960; Trick & Pylyshyn, 1993). Thus, acquiring a verb that describes an event with a high-adicity representation could pose a challenge to the learner if the visual working memory system limits the number of event relations that can be tracked, and thus does not readily yield high-adicity scene percepts. Understanding how children acquire such verbs first requires a concrete understanding of how high-adicity event representations are supported by a visual working memory system with limited capacity.

Importantly, these event representations cannot simply be presumed to be adult-like. Doing so would result in a logical circularity: children's representations of events cannot be presumed to be those under which adults perceive and describe a scene, since the argument structures that adults use to describe an event are precisely what a child is trying to acquire. For this reason, the event representations of children must be probed in a manner independent of language. Previous work that has investigated the participant structure of children's event representations in such language-independent ways suggests that children are able readily view complex events, such as a girl using an instrument to open a box (He, 2015) and a girl taking a truck from a boy (Perkins et al., 2024), under concepts that have three participants.

High-adicity events, such as tradings, provide an ideal case study to probe how developing cognitive abilities may interact with the acquisition of lexical argument structure. In Part 1 of this dissertation, we take a crucial first step towards understanding which scenes children are able to readily represent as having a high number of participants, and probe how they may be able to do so. We approach this question through a series of experiments with both children and adults, in which they viewed various scenes that could plausibly be viewed under a 4-participant representation, and comparing these scenes to scenes in which the participant structure has been manipulated. This work addresses a critical gap of understanding the nonlinguistic representations of such high-adicity concepts in both children and adults.

We begin Part 1 with Chapter 2, in which we introduce the extralinguistic cognitive system that can support these conceptual representations, visual working memory, as well as the potential challenges this developing system can impose on the acquisition of verbs that label high-adicity concepts like TRADE. We then introduce previous linguistic analyses of *trade*, highlighting ways that learners might map between language and conceptual representations when acquiring verbs like *trade*. In Chapter 3, we present findings that adults represent a trading scene under a four-participant TRADING concept in which both traders and both traded items are explicitly represented. By comparing against another plausibly four-participant event, we find converging evidence that adults view this trading scene under a single concept, and not as two sequential GIVINGS. In Chapter 4, we present findings that children aged 3.5-5.5, just like adults, represent all four participants in a trading scene. We also conduct a survey of child-directed speech, which highlights the infrequency of 4-place uses of *trade* (e.g., *Dan and Sue traded a truck and a ball*).

Our findings that preschool-aged children represent all four characters in a trading scene as participants highlight a potential challenge for the acquisition of predicates that label such high-adicity concepts. How might children map between scenes that they represent under 4-place concepts and sentences that may not always have 4 arguments? Theories of how children map between event participants and linguistic arguments, known as linking theories, make different predictions for these predicates. For example, linking theories that require one-to-one mapping of event participants to linguistic arguments (Fisher, 1996; Lidz & Gleitman, 2004; Naigles, 1990; Yuan et al., 2012) may predict additional challenges for learning verbs like *trade*, which can occur with fewer than 4 linguistic arguments. On the other hand, linking theories that rely upon exploiting thematic links between participants and arguments (Baker & Levin, 2015; Dowty, 1991; He, 2015; Jackendoff, 1992; Perkins et al., 2024; Pinker, 1984; Williams, 2015) suggest that children could exploit their knowledge of thematic roles to map between the relevant participants and the linguistic arguments in the sentences they hear. We discuss possible thematic linking strategies for the acquisition of high-adicity verbs in the discussion section of Chapter 4. Overall, these findings lay the foundation for future work investigating the conceptual representations of such high-adicity event types in even younger infants, as well as a more nuanced investigation of the internal structure of these event types. By investigating the non-linguistic conceptual representation for this complex event type, we provide a necessary first step for future theoretical investigation of the types of strategies children could use in learning verbs.

1.2 Extralinguistic cognition in argument structure processing

Once children manage to acquire verb argument structure knowledge, they must still learn how to apply this knowledge in real-time sentence processing, in which they must build structure and extract meaning from sentences they hear in real-time. But still-developing extralinguistic cognitive systems can influence how children deploy their linguistic knowledge in both sentence comprehension and production – potentially even masking linguistic competency through performance limitations (Lidz et al., 2017; Omaki & Lidz, 2015; White & Lidz, 2022). For this reason, diagnosing a child's developing linguistic knowledge requires a detailed understanding of the precise mechanisms by which this knowledge interacts with extralinguistic cognition at a given point in development. Even more critically, developing cognitive systems outside of language influence the type of linguistic input children can learn from. If children fail to parse the sentences that they hear veridically, they may not be able to draw appropriate generalizations on the basis of those sentences (e.g., Lidz, 2023; Lidz et al., 2017; Perkins et al., 2022). Consequently, understanding the precise mechanism(s) by which extralinguistic cognitive systems interact with sentence processing and understanding throughout development is crucial not just because they can influence the behavioral responses researchers rely on to diagnose linguistic competence, but also because the learning process depends on children's linguistic representation of their input, which may not be comparable to an adult's representation.

Extralinguistic cognitive systems play a significant role in sentence processing in both children and adults (Gibson et al., 2000; Hsu et al., 2021; Lewis, 1996; Novick et al., 2005; Trueswell et al., 1999; Woodard et al., 2016, among others). Sentence processing is an incremental and predictive process (Borovsky et al., 2012; Frazier & Rayner, 1982; Gordon & Chafetz, 1990; MacDonald et al., 1994, among many others). We know this from a rich sentence processing literature which has examined adults' incremental parsing decisions in ambiguous and temporarily ambiguous sentences. Children also appear to parse sentences predictively and incrementally (Lidz et al., 2017; Omaki et al., 2014; Snedeker & Trueswell, 2004; Trueswell et al., 1999; White & Lidz, 2022, among others). However, their performance with certain temporarily ambiguous sentences, like (1a), highlights a notable divergence from adult behavior:

(1) a. Put the frog on the napkin in the box. *Temporarily ambiguous*

b. Put the frog that is on the napkin in the box. *Unambiguous*

Unlike adults, who can correctly parse this sentence such that *in the box* is the destination for the frog, 5 year-old children appear to get "stuck" in their initial prediction that the final destination for the frog is the napkin (Trueswell et al., 1999). This effect endures even in the presence of visual referents that guide adults' interpretations (Tanenhaus et al., 1979; Trueswell et al., 1999) or when referential information is pragmatically enriched, such as the experimenter highlighting the initial locations of the frogs or adding prosodic cues (Weighall, 2008). However, children's performance with the unambiguous sentence (1b) is adult-like. Children's interpretation of the temporarily-ambiguous sentence in (1a) appears to arise as a result of the knowledge that *put* is a ditransitive verb, which leads to an early commitment to an analysis where the upcoming prepositional phrase *on the napkin* is the goal argument for the verb (Trueswell et al., 1999). Adults make the same initial commitment (Trueswell et al., 1999). However, adults are able to revise from the misinterpretation that results from this initial commitment and arrive at the correct interpretation. Children, on the other hand, are unable to recover from their early commitment and are therefore garden-pathed more severely.

Two prominent proposals have been put forth to explain this so-called "kindergarten path" effect. In one proposal, working memory limits cap the number of possible parses maintained by the child (Lewis, 1996; Trueswell et al., 1999, among others): at the crucial choice point, only the parse that leads to an incorrect interpretation is maintained, prohibiting later revision even in the face of enhanced referential cues. However, this stands in contrast with more recent findings that it is cognitive control – and, crucially, not working memory – which correlates with children's (Woodard et al., 2016) and adults' (Hsu et al., 2021) ability to revise from incorrect interpretations. Thus, the "kindergarten-path" provides another interesting case-study for the interaction of developing extralinguistic cognition with verb argument structure acquisition. Acquiring verb argument knowledge comes with its own drawbacks for the young language learner: it can lead to incorrect early commitments over the course of sentence processing, from which it may or may not be possible to recover. If these commitments affect the way that children represent their linguistic input, this could ultimately hinder additional language learning (e.g., Lidz et al., 2017).

In Part 2 of this dissertation, we probe the way that early argument structure knowledge, in conjunction with developing memory and inhibitory abilities, can affect children's ability to parse temporarily ambiguous sentences. In Chapter 5, we begin by introducing the literature linking the immaturity of these cognitive systems with children's non-adult like processing behavior, which tends to be insensitive to context (visual or linguistic) and heavily reliant on verb biases. We then present a computational model of sentence processing across development in Chapter 6. In order to probe the effects of developing cognitive control and working memory separately, we developed a model with two individually-adjustable parameters that parallel each extralinguistic cognitive system. By formalizing these systems as individual parameters that can be independently adjusted, we are able to isolate the contribution of each developing system to sentence processing. We evaluated the model on the test sentences from Trueswell et al. (1999) and find that our model can capture the findings from this study: at child-like settings of either parameter, we find that the parser does not find the intended analysis for temporarily-ambiguous sentence (1a), but finds the intended analysis for the unambiguous sentence. At adult-like settings, our parser finds the intended analysis for both sentences. Moreover, our model makes testable predictions for the developmental time-course of an alternative analysis of the temporarily ambiguous sentence (1a) from Trueswell et al. (1999) that has not previously been considered experimentally.

This parser sets the stage for a variety of further computational and experimental work investigating how these two cognitive systems interact with parsing throughout development and into maturity. Generating a clearer understanding of how these systems interact with parsing over development can inform the ways that behavioral evidence from children's sentence understanding is used to diagnose a child's linguistic knowledge at any given point in development. Additionally, this understanding can guide theories of language learning by clarifying the nature of the linguistic representations the learner is learning from at various points in development.

1.3 Dissertation overview

The dissertation is structured as follows. Part 1 begins with Chapter 2, in which we set the stage for the nonlinguistic investigation of children's early verb learning, describing the particular challenges that verb learning poses and the necessity of understanding children's nonlinguistic conceptual representations of scenes. In Chapter 3, we report findings from four experiments with adults investigating the internal structure of their representation of a trading scene. In Chapter 4, we extend these findings to preschool-aged children with a picky puppet experiment. Part 2 of the dissertation pivots away from the question of verb learning towards the question of how verb argument structure knowledge can influence language processing in real time. In Chapter 5, we introduce the relevant literature on children and adults' parsing differences, the architecture and development of two relevant extralinguistic cognitive systems, working memory and cognitive control, and the role(s) of these cognitive systems in real-time sentence processing. Chapter 6 introduces a formal computational model which parameterizes the roles of these two developing cognitive systems.

In sum, this dissertation investigates the interaction of developing extralinguistic cognitive systems with early language learning and processing through the case study of verb argument structure. This question underpins fundamental theories of language learning and use, as the linguistic system does not exist in isolation. At the earliest stages of language development, children are faced with the seemingly daunting task of learning both the meaning and structure of the language used around them. But the extralinguistic cognitive systems supporting this learning are also in their earliest stages of development, potentially placing perceptual and computational constraints on this learning. Similarly, at later stages of development, when children have already learned a significant amount about their target language, the *use* of this acquired knowledge can be influenced and constrained by still-developing cognitive systems. This case study in verb argument structure acquisition provides an example of why we need to study extralinguistic cognition in order to understand the mechanisms that support learning of grammar and meaning, which generalizes to other areas of language development beyond verb argument structure.

Part 1

CHAPTER 2

Verb learning, event perception, and visual working memory

In this chapter, we outline the challenge that verb learning poses to the young learner, discuss aspects of event perception that are crucial for the study of verb learning, and possible perceptual limitations imposed by the visual working memory system on event perception. We focus on *trade* as a case-study in verb acquisition, due to the fact that TRADING is likely to be represented under a 4-place event concept, which could tax visual working memory during development and impede the conceptual-syntax mapping necessary for verb learning. We then discuss possible ways that the young learner could circumvent these perceptual limitations to facilitate the acquisition of *trade*, based on previously proposed semantic analyses of this predicate.

2.1 The challenge of verb learning

Verb learning poses a daunting task over the course of language acquisition. Observation of the external world must play some role in word learning, but this type of observational learning is complicated by the fact that real-world scenes can be uninformative or even misleading: there are many compatible referents that a novel word may be labelling (Quine, 1960). In the case of verb learning, verbs label not only events but the speaker's view of an event (Gleitman, 1990). On many theories, a verb's meaning is relational. As such, the young learner must also identify the relevant relations within the speaker's interpretation of the event. But there are many possible relations that could be relevant: for example, an event in which a dog jumps to catch a frisbee could be described in several ways, each highlighting different relations, as in "The dog jumped", "The dog caught the frisbee", and so on.

The difficulty of identifying the appropriate referent for a novel word is an example of the induction problem of language acquisition (Pinker, 1984, 1989): there are many compatible hypotheses that can explain an observed linguistic pattern. The complicated nature of learning verbs has given rise to a rich literature proposing that infants can exploit their knowledge of systematic regularities between the syntax and semantics of verbs (Dowty, 1991; Fillmore, 1968, 1972; Jackendoff, 1990) to "bootstrap" themselves into verb meanings. Evidence that children implicitly recognize these correlations (Bowerman, 1974, 1977, 1982; Gropen et al., 1989; Maratsos et al., 2014; Pinker, 1989) suggests that they may be able to make predictions about sentence structure on the basis of verb meaning and vice versa. These types of bootstrapping theories, which argue that children map between meaning and structure to learn verbs, typically fall into two camps: semantic (Grimshaw, 1981; Pinker, 1984, 1989) and syntactic (Gleitman, 1990; Landau & Gleitman, 1985) bootstrapping. Semantic bootstrapping theories propose that infants use their conceptual understanding of a speaker's view of an event to make inferences about syntax (proponents of these theories have further argued that semantic bootstrapping can extend to verb argument structure learning; see Pinker, 1989). On the other hand, syntactic bootstrapping theories propose the opposite direction of inference: infants exploit their knowledge of argument structure to identify the relevant participants in a speaker's view of the event that a sentence is describing, in order to learn the meaning of the verb in that sentence. For our purposes, we remain agnostic as to the directionality of bootstrapping and focus instead on children's conceptual representations of events as they pertain to verb learning.

2.1.1 Note on terminology

Prior to further discussing mechanisms and strategies for verb learning, it is necessary to define the relevant terminology used throughout this dissertation. We will use 'argument' to refer to the wholly syntactic notion of a phrase existing in one of the syntactic relations known as subject, direct object, and indirect object relations. Collectively, these will be referred to as 'argument relations'.

‘Participant’ will be used for individuals represented as existing in a ‘participant relation’ to an event (Perkins et al., 2024; Williams, 2015). Importantly, an event concept entails many relations. However, we will consider only those relations that are privileged psychologically and are represented explicitly in the event concept to be ‘participant relations’. We assume that only some of these event relations stand in such a psychologically privileged relation to the event, as these are the relations that are candidates to be arguments in a syntactic clause. The reason for this assumption is that it constrains the learner’s otherwise infinite hypothesis space for possible argument-participant links: although there are an infinite number of entailed event relations, only certain relations are candidates to be arguments of a clause describing that event. For example, in an event that falls under EATING, there will not only be an eater *a* and something *b* that was eaten, but also a location *c* where the eating is taking place and a time *d* when the eating is occurring. The eating may also proceed for a duration *e* and with a manner of movement *f*. Intuitively, only the relations of the eater *a* and the thing eaten *b* are viewed as potential participants in the event of eating. We formalize this by saying that only these two relations are psychologically privileged in such a way that they might be explicitly encoded as parts of the conceptual structure under which the event is represented.

We also denote these relations in terms of the valence, or adicity, of the concept. Somewhat informally, we will represent adicity through the names of event concepts, where each part of the name denotes a relation in the event: the first part (EATING) is the event kind itself, and the next part lists the participant relations that are explicitly represented in that event concept. Thus, a 1-participant event concept is one which contains a single participant relation, as in EATING-AGENT. A 2-participant event concept is one explicitly containing two participant relations, EATING-AGENT-PATIENT. Crucially, all event concepts that share the first part of the name (e.g., EATING-AGENT and EATING-AGENT-PATIENT share the name EATING) entail the same event relations – they only differ in which participant relations are privileged and explicitly represented in the concept. Recall that an EATING could occur in a given location *c*; although this relation is always entailed in an EATING, it does not fill one of the slots in the valence of the concept. The question of which participant relations are encoded in an EATING event is an empirical one – for any event viewed as an EATING, which of the two potential participant relations, *a* and/or *b*, are explicitly represented as participants? For instance, it is unknown whether the thing eaten *b* is explicitly represented as part of the conceptual structure, or whether it is simply entailed along with the other event relations. This question is not answered by observing the argument structure of the verb *eat*: namely, although *eat* can optionally omit its direct object, the patient in such cases is still necessarily entailed. The empirical question that arises is whether the patient is still explicitly represented as a participant in the scene.

Throughout this dissertation, we will refer to an event concept with *n* participants as an *n*-participant event concept. While the verb-learning literature frequently refers to events viewed under 1- or 2-participant representations as simply '1- or 2-participant events', it is important to note that this simplification abstracts away from the conceptual representations formed by an individual upon perceiving an event. Namely, this simplification assumes that all individuals perceive the same event in the same way, and that they view it under an analogous structure to how that event can be described (i.e., with an agent and patient explicitly represented if the sentence contains a subject and an object). For the sake of conciseness, we will also use '*n*-participant event' to mean 'an event conceived of as having *n* participants' throughout this chapter – but this distinction is crucial for the discussion of event perception in Chapters 3 and 4.

2.1.2 Evidence supporting bootstrapping theories

Since their conception, bootstrapping theories have garnered significant empirical support. Naigles (1990) tested whether 2 year-olds could use syntactic information to learn new verbs. Using the preferential-looking paradigm (Golinkoff et al., 1987), children were presented with matching side-by-side videos of two simultaneous actions. In the training phase, while these two events were shown, a loudspeaker played audio recordings of a sentence containing a novel verb (e.g., *gorp*), which under certain hypotheses would be more likely to be compatible with one of the actions in the scenes. One of the two actions consisted of a causal action (a duck pushing a rabbit into an odd bending position) and the other consisted of a non-causal action (the duck and the bunny making the same arm gesture). A sentence with the verb in a transitive frame – *Look! The duck is gorping the bunny!* – is thought to be more compatible with the causal action due to the robust correlation between transitive verbs and causal events (Hopper & Thompson, 1980), a correlation which could guide children to associate these verbs with such events. Children were also familiarized to sentences with the verb in an intransitive frame – *Look! The duck and the bunny are gorping!*. At test, children were presented with two new videos, each depicting one of the actions, and heard the prompts *"Where's gorping?* or *"Find gorping now!"*. Naigles (1990) found that children looked significantly more at the causal action when they had been familiarized with the transitive sentence, indicating that they could use the syntactic frame of the novel verb to draw inferences about which of the two actions the verb labeled. Similar findings have also been found in 15 and 19 month-olds (Jin & Fisher, 2014; Yuan et al., 2012).

Infants have also been shown to use syntactic structure to infer a verb's meaning independent of a visual referent. Yuan and Fisher (2009) familiarized 2 year-olds to sentences with novel verbs occurring either in transitive (*"She blicked the baby!"*) or intransitive (*"She blicked!"*) frames. The children were then shown two side-by-side videos depicting either an event intended to be seen with one participant, or an event intended to be seen with two participants, while hearing the verb in isolation (*"Find blicking!"*). Children who had heard the transitive frame looked reliably more to the plausibly two-participant event than the plausibly one-participant event, indicating that they used the syntactic frame of the novel verb to infer something about verb meaning in the absence of a visual referent during familiarization with the verb. Arunachalam and Waxman (2010) confirmed that infants were using information about the syntactic frame, rather than superficial information about the number of nouns in the familiarization dialogue, by introducing infants to sentences containing two nouns in both the transitive (e.g., *The lady blicked my brother!* and intransitive conditions (e.g., *The lady and my brother blicked!*). This ruled out the possibility that infants were simply looking at the scene with two actors after they heard transitive sentences listing two nouns and the scene with only one actor when only one noun was listed in the intransitive sentences. Once again, infants showed a preference for the plausibly 2-participant event after hearing the sentences in transitive, but not the intransitive, frame.

Arunachalam and Waxman (2010) and Yuan et al. (2012) also found that the children familiarized with intransitive sentences looked equally to both causal and non-causal actions; indeed, this has been a general trend in the literature (for a review, see Noble et al., 2011). Unlike transitive verbs, which almost always describe causal events, intransitive verbs are compatible with both causal and non-causal events (e.g., *The vase broke* could describe a scene in which an agent caused the vase to break, or one where the vase breaks independent of any agent causing it to break). Looking equally at both scene types after being familiarized with intransitive frames indicates that these children accepted both a causal and non-causal scene as possible referents for the intransitive frames. However, it is unknown whether children think that both scenes, perceived as the experimenters intended them, are good referents for an intransitive verb, or whether they instead perceive the '2-participant' scene under a different structure than intended. For example, infants may be viewing the presumed-2-participant scene under several separate '1-participant' concepts. This second hypothesis leaves open the possibility that children think that intransitives must label events viewed as only having one participant, but are ambivalent in these studies due to the fact that these experiments do not explicitly control for their scene percepts. This indeterminacy in children's scene representations highlights the necessity of studying early scene representations in order to diagnose bootstrapping.

2.1.3 Diagnosing scene percepts for bootstrapping

Diagnosing the event representations of infants is necessary for the identification of more precise mechanisms by which bootstrapping is occurring. For example, past work has shown that diagnosing event representations can help decide between two specific proposals for how infants map between structure and meaning. On one proposal, children use one-to-one matching, matching the number of 'argument places' in a sentence to the number of 'participant roles' encoded in their representation of a scene (Fisher, 1996; Lidz & Gleitman, 2004; Naigles, 1990; Yuan et al., 2012, among others). In other words, under this proposal, infants expect a sentence with *n* arguments to describe an event viewed as having exactly *n* participants, for any number *n*. On another proposal, children expect no such one-to-one alignment: instead, they map between the specific contents of their syntactic and conceptual representations through thematic linking (Baker & Levin, 2015; Dowty, 1991; He, 2015; Jackendoff, 1992; Perkins et al., 2024; Pinker, 1984, 1989; Williams, 2015, among others). Here, infants link syntactic positions (such as the 'object of transitive clause') to participant relations (such as 'patient of a causal event').

Perkins et al. (2024) used a scene viewed under a 3-participant concept, TAKING, to adjudicate between these two hypotheses. In this case, one-to-one matching predicts that infants would not accept a 2-argument sentence (e.g., *The girl pimmed the truck*) as a description of a 3-participant event, a scene where a boy is holding a truck, then a girl takes the truck by sliding it towards herself. Thematic linking, on the other hand, predicts that infants will accept this description so long as they represent this event as having a girl as an agent and a truck as a patient. Perkins et al. (2024) found that infants accepted the 2-argument sentence for a TAKING scene. These findings suggest that infants use thematic linking rather than one-to-one matching.

In order to draw this conclusion, Perkins et al. (2024) needed to first establish that infants viewed the taking scene under a 3-participant concept. To do so, they adapted a method inspired by Gordon (2003) and introduced in He (2015) and Wellwood et al. (2015) to probe infants' nonlinguistic representation of a 'taking' scene with human actors. These previous studies asked whether participants would view what was plausibly a change in participant structure – for instance, from an event seen as having 3 participants to one seen as having 2 – as more noteworthy than a change in another physical property of the event, such as manner of motion. The logic in these experiments hinges on the fact that event concepts always entail many event relations (see §2.1.1): for example, if an event is a TAKING, it has an agent of taking, a patient that is taken, and a source from whom the patient was taken. It also has a manner of taking, the duration of taking, the location in which the taking occurred, and so on. It is an empirical question whether the agent, patient, and source are explicitly represented as participant relations in the conceptual str ucture under which someone is viewing a particular scene. All else being held equal, if the viewer treats a change in one of these hypothesized participant relations as more noteworthy than a change to another event relation, such as manner of motion, this may be taken as evidence that this hypothesized participant relation is explicitly represented in the conceptual structure under which they viewed the event. Otherwise, it may be possible that one of these relations (for example, the source), has the same status as the other entailed non-participant relations (such as manner, location or duration), and is not being explicitly represented as a participant.

Perkins et al. (2024) familiarized 10-12 month-old infants to videos of a girl picking up a truck, with a boy sitting idly by, and then compared the relative dishabituation to two scenes: in one scene, the girl was now taking the truck from the boy, who was holding the truck. In the other, the girl now slid the truck towards herself instead of picking it up, while the boy continued to sit idly by. They found that infants viewed the change to the boy's involvement as more noteworthy than a change to the manner in which the truck was moved. Ensuring that no other perceptual differences could explain this pattern of results, this suggests that infants viewed this participant change as an important conceptual difference: infants viewed the TAKING scene, but not the 'PICKING-UP' scene, under an event concept in which the girl, truck, and boy were all explicitly represented as participants.

By explicitly pitting changes in an entailed non-participant event relation (manner of motion) against changes in participant relations, He (2015), Perkins et al. (2024), and Wellwood et al. (2015) provide a valuable diagnostic tool for event representations. Moreover, Perkins et al. (2024) is one of the first to explore early verb learning with a high-adicity event concept, TAKING. The bootstrapping literature has focused almost exclusively on basic transitive and intransitive sentences corresponding to 1- to 2-participant event concepts. But children must also learn verbs that have more complex argument structures, such as those that encode relations between three and four participants. Perkins et al. (2024) raises the question of what other high-adicity concepts are possible candidates for verb meanings.

2.2 High-adicity event concepts in development

Event concepts with high adicity, such as 3- or 4-place concepts, are rarely lexicalized (Pietroski, 2010), and tend to only be limited to predicates of exchange like *give, take, buy, sell,* and *trade*. These event types hold special status in human cognition even early in development, likely due to their social importance, inviting the question of how children acquire verbs that label these event types. Humans may be special among primates in the prevalence of givings and tradings: of the primates, humans are one of only a few primate species that frequently engage in 'active resource transfers' (or givings) both within and outside of family units, while other non-human primates typically engage in 'passive resource transfers' (or takings), and only between related individuals (Brosnan & De Waal, 2002; De Waal, 1989; Feistner & McGrew, 1989; Ueno & Matsuzawa, 2004). The active giving of resources has been recorded even in the earliest of human societies (Enloe, 2003; Gurven, 2004; Jaeggi et al., 2010).

Giving events are also perceived with special status early in human development. Schöppner et al. (2006) tested whether infants aged 9-12 months can detect role reversals in 3-place GIVINGS. Infants were habituated to scenes in which one puppet gives a flower to another puppet. Upon habituation, infants were shown 6 test scenes, in which the puppets switched spots spatially (i.e., the puppet originally on the right now appeared on the left and vice versa). Half of these test trials consisted of a role reversal, where the puppet that previously received the flower was now giving the flower to the other puppet. The other half of the test trials consisted of a direction reversal (by virtue of the fact that the puppets changed positions): the transfer occurred in the opposite direction of motion, but with the giver and receiver roles remaining consistent with the habituation phase. They found that both 10.5 and 12 month-olds noticed the role reversal reliably more than the change in direction, while 9 month-olds dishabituated to both scene types equally. Thus, by 10.5 months, infants appear to notice changes to the relevant relations in a giving event above and beyond changes to other aspects of the scene – suggesting that at this age they privilege the participant roles over other properties.

Infants also appear to have early expectations about the minimum number of participants required for giving events. Tatone et al. (2015) conducted a series of experiments to probe 12 month-old infants' representations of giving and taking scenes and found that infants notice the difference between 'giving' and 'taking' actions. However, they did not appear to always notice when the third party in a 'taking' scene changed, while noticing changes to the identity of the third party in a 'giving' scene. As discussed previously, Perkins et al. (2024) found that infants do readily represent the Source in certain TAKING scenes. The discrepancy in these findings may be due to the fact that a variety of social cues can alter infants' and adults' representations of similar scenes (Hafri et al., 2013; Papeo et al., 2017). Indeed, eye contact between the same types of shape-based characters as in Tatone et al. (2015), which occurred before movement of the patient was initiated, led infants to represent the Source (Tatone & Csibra, 2020). Alternatively, it is possible that the infants in Tatone et al. (2015) may have viewed the scenes that were tested as an OBTAINING, which does not have a source, rather than as the intended TAKING. Because this wasn't tested explicitly, there is no way of telling under which concept these infants viewed the scene. Thus, it is difficult to draw a meaningful conclusion from the lack of SOURCE representation in Tatone et al. (2015).

Infants also appear to privilege giving actions over taking actions with respect to encoding the reciprocity of these resource transfers. Tatone and Csibra (2020) tested 12 month-olds in a series of experiments with kinematically identical giving and taking actions. They found that infants looked longer to a reciprocated 'giving' than to reciprocated 'taking', suggesting that the reciprocity of resource transfers was only encoded for giving actions, and not taking actions. This is particularly interesting in light of the fact that two reciprocal GIVINGs could be viewed under a larger TRADING concept.

In summary, early in development, infants privilege giving, taking, and trading event types. These high-adicity, socially-important event types are likely candidates for verb meanings. We know from Perkins et al. (2024) that infants are readily able to represent a TAKING scene under a 3-participant concept. Moreover, Perkins et al. (2024) found that infants accept a 2-argument clause as a label for the scene, suggesting that children rely on thematic linking, and not one-to-one matching of arguments and participants, to learn the meaning of new verbs.

In this portion of the dissertation, we investigate *trade* as another case study for high-adicity verb learning. *Trade* (along with *buy* and *sell*) is comparatively unique in that there do not appear to be other lexicalized verbs with more arguments across the world's languages, excluding cases of verb serialization. Serial verb constructions are those in which multiple verbs are combined and act as a single syntactic unit (2):

(2) *ó* *tì-wà-rà* *étéré* *à*
he hit-split.open-TENSE plate the
"He shattered the plate" (Igbo; Lord 1975:27)

However, such serializations are not examples of simple lexical entries in the same way as *trade*. *Trade* is further valuable as a case study due to the fact that TRADINGs are 4-place event concepts, containing two Agents and two Items-Traded. Interestingly, these two sets of participants stand in the same relation to the TRADING event concept, providing an interesting new avenue of investigation for theories of thematic linking and bootstrapping. Because acquiring a high-adicity verb like *trade* requires mapping a linguistic form onto a conceptual representation of an event under the TRADING category, young learners must be able to perceive events in the world as 4-place events. Even if bootstrapping from syntax could provide some guidance for the young learners (discussed further in §2.2.2), we might have reason to question the perceptual support for high-adicity events like this, as we now discuss.

2.2.1 Perceptual support for participant relations in high-adicity events

In adults, recognition of event relations occurs automatically and rapidly in visual perception, indicating that the human perceptual system is tuned to extract event relations from brief visual exposure. Hafri et al. (2013) conducted a series of experiments testing adults' ability to extract information from a still photograph of an event after short (73ms) and shorter (37ms) exposures.

They found that the participants could reliably identify the event type ("Did you see kicking?"), the Agent or Patient status of the actors ("The girl acted on the boy" or "The boy was acted upon by the girl"), as well as the veracity of a sentence describing the event ("The girl kicked the boy") even after the shorter exposure, indicating that adults are able to extract a host of relevant information about an event from brief visual exposure. The extraction of event roles thus appears to occur in only a fraction of the time it takes for facial and bodily emotions to be registered (115 ms; Meeren et al., 2005) or for neural responses specifically attuned to the recognition of bodily-shapes to be registered (190 ms; Pourtois et al., 2007).

In a different task, Hafri et al. (2018) tested the response time of participants asked to track the location of a particular actor (male/female or red-/blue-shirted) in various sequential scenes, pressing buttons to indicate the relevant actor appearing on the left or right side of the screen. They found that reaction times were increased whenever the relevant actor's role in events was changed from trial to trial – in other words, when the actor's role changed from Agent to Patient or vice versa. Thus, although the actor's event role was orthogonal to recognizing the identifying characteristics of that actor, the alteration of an actor's role delayed response time, indicating that event relations are extracted automatically, even when they are not being explicitly probed.

This rapid extraction of event relations in visual perception is likely supported by the object file system, a component of visual working memory. In an object file system, individuals are represented implicitly, as "object-files" that can be evaluated through one-to-one correspondences (Simon, 1997; Uller et al., 1999) or through attentional indices to objects in an array (Feigenson & Carey, 2003; Feigenson et al., 2002; Leslie et al., 1998). There is evidence that event-roles are neurally represented as such object-files in working memory: Yu et al. (2024) compared neural activations for images consisting of coordination (a lion and an elephant standing neutrally) or event relations (a lion pushing/hitting/pointing at an elephant) and found a greater sustained ERP response for only the images containing relations. This sustained response, in an area associated with working memory, indicates that the object file system may have a specialized neural pathway for the encoding of event relations but not coordinations.

Constraints on the object file system, however, highlight a possible perceptual constraint on the mapping required to acquire verbs like *trade*: throughout development, the visual working memory system has a limit of 3-4 items. Adults are typically able to identify and track up to 4 objects within their visual working memory at a time. Sperling (1960) first investigated the amount of information that can be extracted upon a brief visual exposure. The stimuli consisted of sets of letters, either grouped or spaced apart, ranging in length from 3 to 6 letters. He found that participants could typically only remember on average 4 of the letters. This limit of 4 occurs in multiple other studies investigating the ability of adults to track and encode items visually. Pylyshyn and Storm (1988) tested participants on their ability to fixate multiple moving objects at once, and found that participants could track at most 4 items in parallel, without sequential eye fixations. This finding has since been replicated in a variety of studies investigating parallel object tracking and enumeration (Scholl & Pylyshyn, 1999; Trick & Pylyshyn, 1993). Taken together, these findings suggest that adults can maintain at most 4 items at a time in their visual working memories. As such, although tracking the participant relations in a 3 or even 4 participant event might not *surpass* the visual working memory limit of adults, it may place the visual working memory system at its limit.

Infants have even more stringent caps on visual working memory at the ages they are most actively learning verbs. To probe the visual working memory limit of children, Feigenson and Carey (2003) developed a paradigm in which 14 month-old infants watch an experimenter place a set number of toy balls into a box. The experimenter then surreptitiously removes one (or more) of the balls. The amount of time the child spent searching for the ball, compared to a baseline condition in which no balls were removed, was used as a measure of children's memory for the number of balls placed in the box. Although children were able to track up to three balls being placed in a box, when the number of balls exceeded three, children no longer searched for the additional balls. This suggests that children could remember the presence of at most three balls at a time, indicating a visual working memory limit of 3.

These visual memory limits are not insurmountable, however – certain strategies allow both adults and children to circumvent the limit. One such strategy involves "chunking" items that are conceptually or perceptually similar (Chase & Simon, 1973; Feigenson & Halberda, 2004, 2008; Gobet et al., 2001; Miller, 1956; Moher et al., 2012), hierarchically reorganizing them and thereby increasing the number of items that can be maintained in working memory at once. For example, compared to a random 9-letter string like SHKIUEJWP, it is easier to remember the 9-letter string PBSCNNBBC if the viewer realizes they can divide it into three smaller units comprised of television acronyms: PBS, CNN, BBC. Even children are able to chunk items to overcome this limit: Feigenson and Halberda (2008) tested 14 month-olds and found that they continued to search for four items from two conceptual groups (2 cats, 2 cars), but did not do so for four items belonging to the same conceptual group (4 toy cats). These children were also able to track four identical items if the items were given two different labels (2 *daxes*, 2 *blickets*), suggesting that language can help "chunk" items in conceptual representations, even if they are perceptually identical.

Infants are even able to form conceptual chunks around possession and ownership: using the same manual search paradigm, Stahl et al. (2023) found that infants could use cues of ownership to remember four perceptually identically items by organizing them into chunks by possessor. Infants were tested on sets of 4 identical blocks, which should exceed their memory capacity in the absence of any chunking strategy. Possession was marked by an experimenter using a stuffed animal to move the blocks, saying "I'm going to put mine here!" each time they manipulated a block. When the items were all "possessed" by the same stuffed animal, children did not continue their search after 3 items were removed, indicating they could not maintain the fourth block in their visual working memory. However, if two different animals each "possessed" two of the blocks, the infants searched for all 4 blocks. These results were not replicated when the blocks were "possessed" by cups matching the color and patterns of the stuffed animals. Thus, possession is a relevant conceptual category that can be used for chunking groups of perceptually identical items, but is only applicable for possessors perceived as animate.

The ability to deploy a chunking strategy – especially by possession – could prove particularly helpful in the acquisition of *trade* by allowing the young learner to circumvent visual working memory limitations and track all four participant relations. Namely, if the young learner hierarchically reorganizes the relations in their representation of a trading event into smaller chunks, each with fewer participant relations, this could facilitate acquisition of the verb *trade*. What might this reorganization look like? The meaning of this predicate can be analyzed decompositionally in several ways, each of which could have analogues in different perceptual chunking strategies.

2.2.2 Semantic analyses of predicates like *trade*

Jackendoff (1992) analyzed predicates of exchange like *buy, pay, sell,* and *trade* as two related transfers of possession. For example, in his analysis, the conceptual representation (3) for *buy* includes both information of a transfer and a countertransfer:

(3) X buy Y from Z for W.

a. GO_{poss} ([Y], [FROM [Z] TO [X]])

b. GO_{poss} ([W MONEY], [FROM [X] TO [Z]])

where an individual Z transfers Y to individual X, and where X transfers some amount W of money to Z in exchange. These events are combined into a single lexical entry (4) through the use of the modifier EXCH, which functions to foreground the relevant transfer compared to the countertransfer (c.f. *pay*, which foregrounds the countertransfer of money), where *i* is the index for the subject:

(4)

$$\begin{bmatrix} \text{buy} \\ \text{V} \\ \_\_\_\text{NP}_j \; \langle \text{ from NP}_k \rangle \; \langle \text{ for NP}_m \rangle \\ \begin{bmatrix} \text{GO}_{\text{poss}} \text{ ([ ]}_j\text{, [FROM[ ]}_k^{\alpha}\text{, TO[ ]}_i^{\beta}\text{])} \\ \text{EXCH [GO}_{\text{poss}} \text{ ([MONEY]}_m\text{, [FROM[}\beta\text{], TO[}\alpha\text{]]))]} \end{bmatrix} \end{bmatrix}$$

Extending this analysis, a *trade* could be represented as two related transfers of possession (5):

(5)
$$\left[\begin{array}{l} \text{trade} \\ \text{V} \\ \underline{}\text{NP}_j \,\langle\, \text{NP}_k \,\rangle \,\langle\, \text{for NP}_l \,\rangle \\ \left[\begin{array}{l} \text{GO}_{\text{poss}}\ ([\ \ ]_l,\ [\text{FROM}[\ \ ]_j^{\alpha},\ \text{TO}[\ \ ]_i^{\beta}]) \\ \text{GO}_{\text{poss}}\ ([\ \ ]_k,\ [\text{FROM}[\beta],\ \text{TO}[\alpha]]) \end{array}\right] \end{array}\right]$$

However, unlike *buy*, in *trade* it is possible that neither transfer nor countertransfer is foregrounded. For this reason, our posited lexical entry does not have the modifier EXCH subordinating the countertransfer. Note that (5) is a decomposed lexical entry for the verb and contains both a hierarchical conceptual structure and a formalized linking between this structure to the syntax of the sentence in which the verb will appear. Going forward, we will abstract away from the linguistic content of these representations, as we are primarily interested in nonlinguistic conceptual structure. We denote this nonlinguistic representation using concise notation as in (6), which contains the same information as (5) without the additional linguistic content.

(6) a. **GIVING1**-AGENT-RECIPIENT-ITEM

b. **GIVING2**-AGENT-RECIPIENT-ITEM

In using the notation in (6) we abstract away from further formal details of the structure under which events are represented in the semantics. Following Perkins et al. (2024), we use the notation in (6) as a way of denoting the valence and participant relations of the event representation, and the way that they may be hierarchically organized within the larger event concept.

The chunked representation of a TRADE in (6) would consist of a hierarchical reorganization of the relevant relations into two GIVINGs, each consisting of three participants: an Agent, a Recipient, and an Item. Such a perceptual representation would not exceed the visual working memory limit during development, because infants are able to maintain up to 3 items within their visual working memory. Thus, by virtue of having only three participants each, chunking the TRADE into two GIVINGS with representations that do not exceed this limit could help infants track all of the relevant participants, thereby facilitating the acquisition of *trade*.

Another semantic analysis of this predicate could find a different parallel in perceptual chunking. In this analysis, the 4-place meaning of *trade* is derivationally related to its reciprocal collective predicate. A collective predicate consists of referents in a conjoined subject (e.g., *X and Y*) mapping onto the same kind of participant relation within an event. Collective reciprocal predicates, like (7b), often exist in alternation with symmetric binary predicates (7a). Symmetric predicates are those predicates in which, for every *x* and *y*, the statement *R(x,y)* is logically equivalent to *R(y,x)* (Gleitman et al., 1996; Winter, 2018).

(7) a. Sue dated Dan ⇔ Dan dated Sue / Mary is John's cousin ⇔ John is Mary's cousin

b. Sue and Dan dated / Mary and John are cousins

The observation that reciprocal collective predicates frequently have systematic morphosyntactic alternations with symmetric predicates has led to the conclusion that the meanings of symmetric predicates and collectives are importantly related. The direction of the derivational relationship is debated, with some arguing that the collective form of a predicate is derived from the binary symmetric form of that predicate (Gleitman et al., 1996) and others arguing the opposite directionality (Winter, 2018). However, regardless of the direction of the relationship between symmetric predicates and reciprocal collectives, the mapping of individual referents within a predicate to the relevant participant relation within an event remains constant (Nordlinger, 2023). For example, in (7b) both Sue and Dan stand in the same relation to *dating*; the same holds in (7a).

Trade also exhibits the behavior of a symmetric predicate, because if *Sue traded with Dan* is true, *Dan traded with Sue* is necessarily true as well, following the pattern *R(x,y)* ⇔ *R(y,x)*. In the case of *trade*, we observe that typical collective-binary symmetric alternation may be scaled up to an alternation between a binary collective and a 4-place symmetric predicate. If (8a) is true, the binary collective (8b) is also true:

(8) a. Sue traded Dan a ball for a truck ⇔ Dan traded Sue a truck for a ball.

b. Sue and Dan traded a ball and a truck.

What is the sense in which (8a) is symmetric? It seems that more structure is needed within the predicate for this logical property to hold. Two possibilities for this internal structure, where the arguments are organized into subgroups in different ways, are shown in (9):

(9) a. *Trade((s,d),(b,t))* ⇔ *Trade((d,s), (t,b))*

b. *Trade((s,b), (d,t))* ⇔ *Trade((d,t), (s,b))*

Notably, the type of symmetry in (9a) differs from the cases previously examined: represented with this structure, *trade* appears to be symmetric only in so far as the elements within the inner brackets are each symmetric, even as the elements within the outer brackets are not[1]. This type of 'second-order' symmetry differs from a 'first-order' symmetry that could also apply to *trade*, as in (9b), in which the elements are grouped by initial possession (i.e., Sue with the ball, Dan with the truck). Regardless of the exact grouping of elements, it appears that the 4-place use of *trade* in (8) exists in alternation with a binary reciprocal collective[2]. Thus, it may be taken as a generalization of the binary symmetric-unary collective alternation (Winter, 2018). Although *trade* can be used in reciprocal constructions in English, it is unknown whether *trade* is inherently reciprocal in other languages; future cross-linguistic analysis could reveal whether there is morphological evidence of the reciprocality of *trade* in languages outside of English.

---

[1] Note, however, that these elements are not 'independently symmetric': *Sue traded Dan a ball for a truck* is not equivalent to *Dan traded Sue a ball for a truck*, suggesting that initial possession is reflected in the relative ordering of these arguments.

[2] We also observe that (8b) appears to collapse the non-symmetric aspects (who traded what) of the sentences in (8a); when the subject is collective, it feels more natural for the object to also be collective (*?Sue and Dan traded a ball for a truck*). We leave a detailed analysis of this phenomenon for future investigation.

The analysis that the collective and symmetric predicates are related derivationally might in turn correspond to a different 'chunked' representation of *trade*: one in which the trading scene is viewed as a collective action along the lines of (8b) and (9a), in which the participants are grouped into two chunks, one including the traders and one including the items. This hierarchical reorganization also circumvents the perceptual limits of visual working memory: these two chunks each encode only two participants. Thus, the conceptual representation would be akin to (10):

(10) **TRADE**-[AGENT1-AGENT2]-[ITEM1-ITEM2]

Note that here, TRADE is represented under a single concept; the participants are simply chunked into groups (denoted by brackets) of two participants each. In another possibility, *trade* could be 'chunked' in the style of 'first-order' symmetry discussed above (9b), with participants chunked by possession, as in (11):

(11) **TRADE**-[AGENT1-ITEM1]-[AGENT2-ITEM2]

Just as before, here TRADE is represented under a single concept, but now the participants are chunked by ownership. This representation may have empirical support from Stahl et al. (2023), which demonstrated that infants are able to chunk perceptually identical items solely on the basis of possessor (necessarily encoding the possessor with the 'possessee', as this distinguishes the two chunks of items).

In summary, the semantic analyses of predicates like *trade* discussed in this section hold parallels to three possible chunking strategies that could be applied to forming an event representation of a TRADING event. Importantly, all of these representations consist of hierarchically rearranging the participants in an event such that there are fewer than 4 participants encoded within each 'chunk'. Per Jackendoff (1992), a TRADE could be viewed as two causally-related GIVINGS, with one agent giving an item to a recipient, who then reciprocates the transfer with another item. Per analyses of collective and symmetric predicates as in Gleitman et al. (1996) and Winter (2018), a

TRADE could instead be viewed under a single representation, either with the actors grouped collectively and the items grouped as well or with the actors and items grouped by initial ownership. By organizing the actors and items in two chunks each containing only two participants, this could also bypass the perceptual limit of 3 on visual working memory throughout development.

The parallels between the semantic analyses of predicates like *trade* and possible chunking strategies for this event type raise an interesting question for verb learning: how might the kinds of conceptual representations that are readily yielded by the visual working memory system support different types of bootstrapping strategies? For example, if a young learner viewed a trading scene under a representation analogous to a reciprocal collective predicate, they may have expectations about the way this representation relates to clausal syntax: perhaps a verb labelling this event could participate in these kinds of reciprocal-symmetric alternations. However, the representation of a trading scene which consists of two related GIVINGs may not make the same prediction. For this reason, examining chunking strategies may help us determine what bootstrapping strategies could be using when learning a verb like *trade*. We return to this discussion in Chapter 4.

2.3 Summary

Given the semantic complexity of verbs describing events viewed as having many participants, like *trade*, this raises questions of how the young learner may bootstrap themselves into the meaning of these predicates. Currently, significant empirical support for bootstrapping has accumulated for 1- and 2-place predicates; however, very little work has addressed more complex event types. These event types, such as TRADING, are independently interesting from a cognitive perspective given their privileged status in human cognition, both in their ubiquity and the fact that they are among the only concepts with high adicity that are lexicalized as simple monomorphemic verbs. Moreover, as a result of this high adicity, these events have the potential to tax the visual working memory limit over the course of development.

Consequently, understanding how children can acquire a verb like *trade* first requires us to identify the conceptual representations of this event type. Fully understanding the nature of the correspondence between linguistic structure and conceptual representation requires a detailed understanding of not only linguistic representations, but also conceptual representations of events. To our knowledge, no previous study has explicitly examined the *non-linguistic* conceptual structure of TRADING event representations in either children or adults. Thus, we first need to determine which event representations are readily yielded by the perceptual/conceptual system. In the next two chapters, we probe the conceptual representations of a trading scene. In Chapter 3, we describe findings from four experiments with adults investigating whether adults encode all four participants – two actors and two toys – involved in the trading scene, and whether this is accomplished through a chunking strategy analogous to Jackendoff (1992)'s semantic analysis of TRADE as two causally-related GIVINGs. If this chunking strategy is not used, it may follow that one of the other chunking strategies, for example, grouping traders and items-traded, may be a plausible alternative. In Chapter 4, we describe findings from an experiment with preschool-aged children testing their perception of the same trading scene. By testing the non-linguistic event representations of adults and young children, we set the stage for future work with infants investigating the perception of these events at the age they are most actively acquiring verbs.

CHAPTER 3

Perception of complex event types in adults

In this chapter, we set the stage for later investigation of event perception in young children through a series of experiments with adults. In Experiment 1, we investigated the conceptual structure under which adults view a scene depicting a trade, and find that they explicitly represent all four event participants. This type of four-place representation may place the working memory system at its limit: typically, adults can only maintain 3-4 items in visual working memory (Cowan, 2001; Halberda et al., 2006; Scholl & Pylyshyn, 1999; Sperling, 1960; Trick & Pylyshyn, 1993). We then asked whether adults are circumventing this limit by deploying a chunking strategy (Chase & Simon, 1973; Feigenson & Carey, 2005; Feigenson & Halberda, 2004; Gobet et al., 2001; Miller, 1956; Moher et al., 2012), effectively lowering the visual working memory load by chunking the scene into two giving events with three participants each. Experiments 2 through 4 answer this question by pitting our trading scene against a control stimulus which may also plausibly be viewed under a 4-place concept. We find that adults also represent all four participants in this control stimulus, but through a different event structure: this control scene is perceived as two separate but sequential events, but the trading scene is not. A schematic of the experimental design for this study can be found in Figure (3.1).

3.1 Experiment 1

Our first experiment diagnosed the adicity of adults' event percepts through a similarity-rating task that adapted a method inspired by Gordon (2003) and introduced in He (2015), Perkins et al. (2024), and Wellwood et al. (2015). These previous studies asked whether participants would view

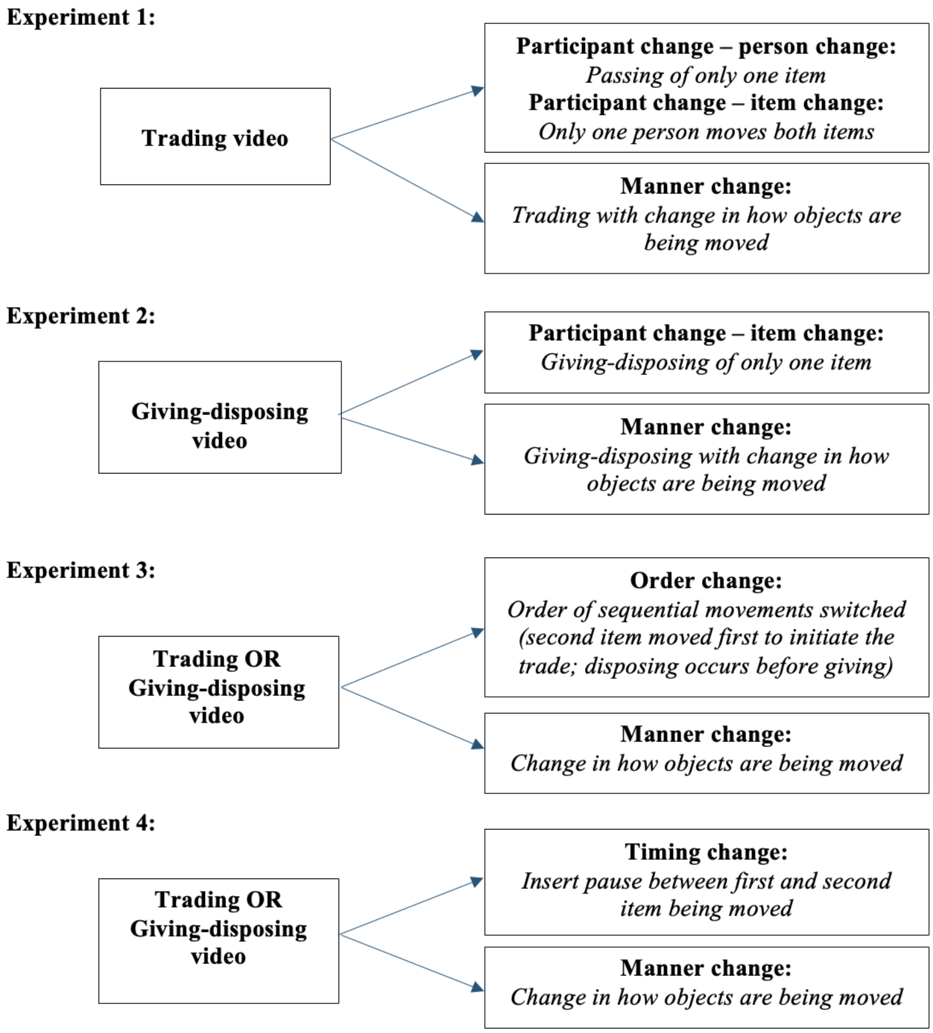


Figure 3.1: Experimental design for Study 1

what was plausibly a change in participant structure – for instance, from an event seen as having 3 participants to one seen as having 2 – as more noteworthy than a change in another physical property of the event, such as manner of motion. The logic in these experiments is that with all else being held equal, if the viewer treats a change in these participant relations as more noteworthy than a change to another event relation, that difference in noteworthiness may be taken as evidence about the conceptual structure under which they viewed the event.

Following this logic, Perkins et al. (2024) used a habituation-switch paradigm to test whether 10-12 month old infants noticed a change in hypothesized participant structure above and beyond a change in the manner of motion (for more detail on this experiment, see Chapter 2, §2.2). Here, we adapted this design to test adults' representations of a scene in which a boy gives a girl a truck, and the girl gives the boy a ball (Figure 3.2).

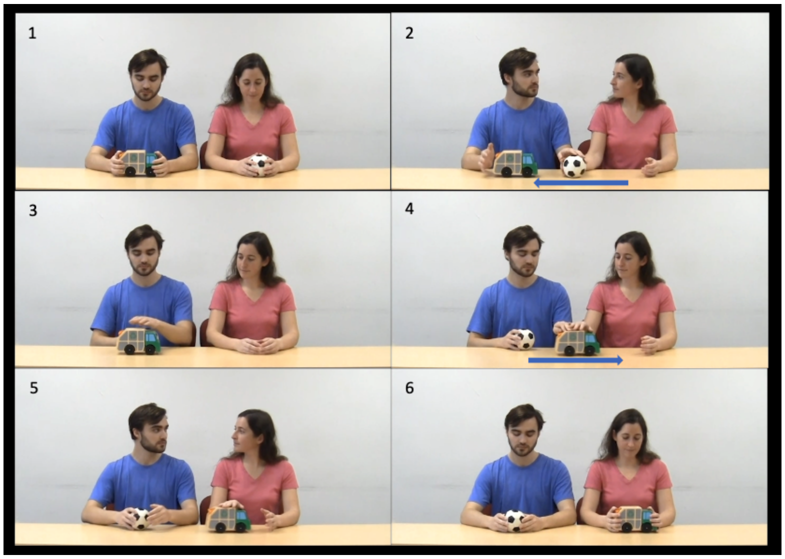


Figure 3.2: Illustration of 'trading'

To do so, we asked adults to compare this video to videos in which one of the actors or items was no longer a participant, or the manner of motion was changed. Under one hypothesis, this stimulus scene might be viewed as a TRADING in which all 4 participants are explicitly represented in the conceptual structure (12):

(12) **TRADING**-AGENT1-AGENT2-ITEM1-ITEM2

If one of the items is removed from the action – for example, the ball is passed back and forth between the actors and the truck remains unmoved – this might now be viewed as a 3-participant PASSING (13). Similarly, if one of the actors looks off to the side while the second actor swaps the two items, this might be viewed as a 3-participant SWAPPING (14):

(13) **PASSING**-AGENT1-AGENT2-ITEM

(14) **SWAPPING**-AGENT-ITEM1-ITEM2

We contrast these changes in participant structure to a change in manner of motion. If the ball and truck are moved not by sliding them on the table, but by lifting them off of the table, this might be seen as a TRADING with a different manner, but not a different number of participants (15). Note that we name this LIFTINGTRADE simply to highlight the difference between this representation with a different manner of motion and the previous TRADING representation, and do not postulate that it is viewed under an entirely different event concept type. Crucially, by virtue of being a manner change, LIFTINGTRADE does not change the adicity of the original conceptual representation.

(15) **LIFTINGTRADE**-AGENT1-AGENT2-ITEM1-ITEM2

If people view the trading but not the passing or swapping scenes under a 4-participant concept, then changing from (12) to either (13) or (14) will involve a change in conceptual structure. All else equal, we predict that these ‘participant changes’ will therefore be viewed as more noteworthy than a manner change. In particular, we would expect changes to the hypothesized event participants to be rated as less similar to the original trading scene than a change to the manner of motion.

3.1.1 Methods

3.1.1.1 Participants

24 adults (12 female; ages 18-63) were recruited via Prolific. Participants were from the United States or the United Kingdom and were paid $6 for participating.

3.1.1.2 Stimuli

We developed a series of video stimuli that manipulate the possible participant structure under which the scenes could be represented. All of the videos contained the same four potential participants: a girl, a boy, a ball, and a truck, all visible throughout the event. The videos always begin with the girl holding the ball and the boy holding the truck, looking down at the respective item unless otherwise specified. In ‘trading’ videos, they exchange the two items, one after the other, after making brief eye contact (Figure 3.2). The eye contact was included to facilitate recognition of the scene as a ‘trading’ by signalling the cooperative and reciprocal nature of the event (for further discussion of the role of eye contact in event representations, see Tatone and Csibra, 2020 and Perkins et al., 2024).

In ‘item-removed’ videos, the event proceeds exactly as in a trading video, except that one of the items is no longer moved. The actors pass either the truck or the ball back and forth, with the second item present but unmoved (Figure 3.3). At the end of the event, the actors return to holding their original items. We made the choice for the second actor to begin and end the scene holding their item for two reasons. First, having the actors hold their respective items allows for the initial and final frames of the video to be consistent across conditions. This is especially important given that the boundaries of events (that is, the beginnings or endings) are the most salient and changes are most likely to be noticed (Baker & Levin, 2015) at these boundaries. Second, leaving the unmoved item completely untouched could reduce its saliency to such an extent that it may not even be noticed by the viewer. However, we want the item to be salient enough as to be a possible candidate participant in the event – if it is still salient, but *not* represented under the event concept, we would expect the viewer to perceive the change as noteworthy.

In ‘person-removed’ videos, the event proceeds exactly as in a trading video except that the items are exchanged by only one of the actors; the second actor looks off to the side and does not participate (Figure 3.4). Together, the person-removed and item-removed videos comprised the participant change manipulations. In ‘manner change’ videos, the event proceeds exactly as in a trading video, but with the actors lifting the items rather than sliding them across the table.

Three tokens of each of the trades, manner changes, and participant-removals were recorded, with the timing of each motion identical across scene types in order to match their perceptual properties. This was achieved by using pre-recorded audio cues and a metronome during filming.

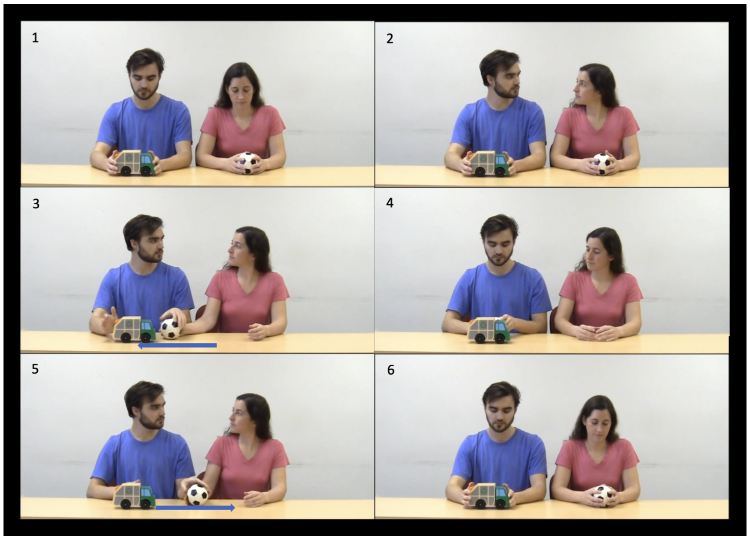


Figure 3.3: Illustration of ‘item-removed ’ video

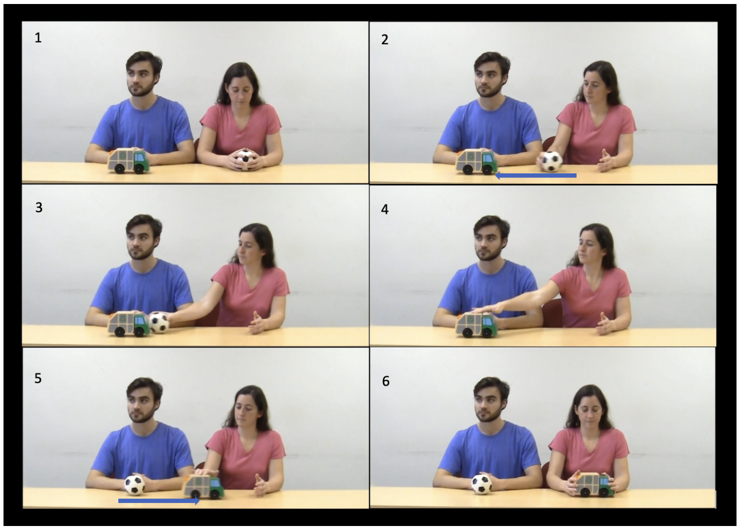


Figure 3.4: Illustration of ‘person-removed’ video

The audio cues consisted of action cues, occurring at fixed intervals, marking when the actors should make eye contact, move the first item, and so on. Each token had a duration of exactly 10 seconds.

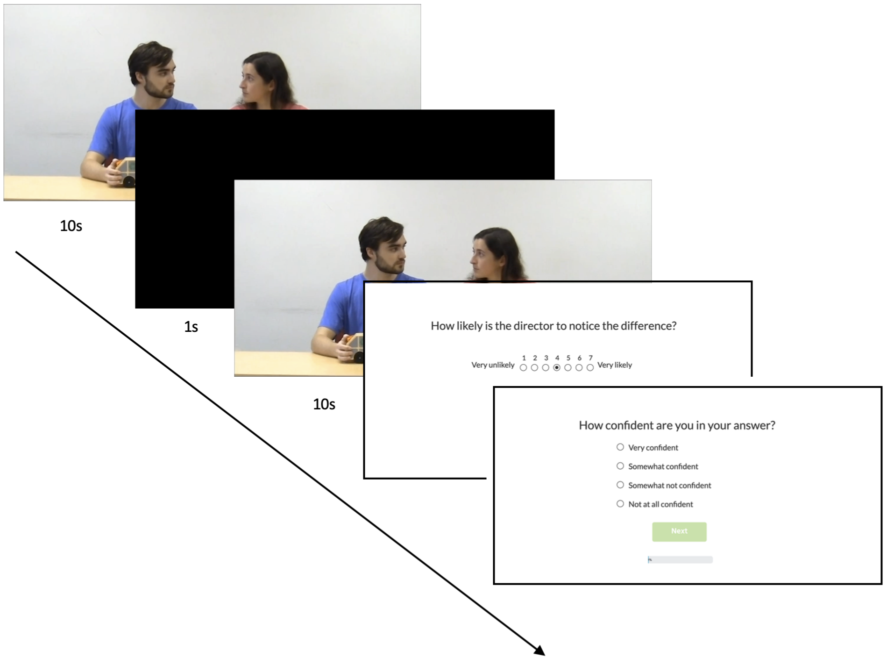
Test trials were created by pairing two tokens together with one second of black screen between them. Each trial was 21 seconds long. Each of the three tokens per event type was paired with all three tokens of the relevant manipulation (for example, Trade Token 1 was paired with Ball-Subtraction Token 1, 2, and 3, etc.). This resulted in 18 total videos for each type of pairing.

Tokens were matched for order of movement within the pair, with either the ball moving first in both videos or the truck moving first, but never a mixture of the two. Baseline control stimuli were developed by pairing two ‘trading’ tokens together; however, no two identical tokens were paired together, leading to a total of 12 control trials (6 trials for each direction of movement). Overall, 120 trials were created for Experiment 1. These trials were broken into two lists, with 66 trials per list: half of the experimental trials (54), and all 12 control trials, counterbalanced for order of presentation and direction of movement.

3.1.1.3 Procedure

The experiment was conducted online through LabVanced (Finger et al., 2017). Participants were told that a video editor had lost footage for a film, and that they needed to judge how likely the film director would be to notice the change between the first video (the “lost footage”) and the second video (a “substitute take”). Videos played automatically and could not be paused or replayed. All trials consisted of a video pair, followed by a 7-point Likert scale with the prompt “How likely is the director to notice the change?”, with 1 being “Very unlikely” and 7 being “Very likely.” Participants were also asked to rate their confidence on a 4-point scale (Figure 3.5). Prior to the test trials, participants were given two practice trials, one with a manner change and one without. No feedback was given throughout the course of the experiment.

Participants were randomly assigned to one of the two lists, with condition manipulated within participants. Condition type was pseudo-randomized by trial such that no condition was seen more than twice in a row. Test trials were counterbalanced across participants for both the order of token-type presentation (trade-change vs. change-trade) and the order of item movement (ball-first vs. truck-first).



How likely is the director to notice the difference?

Very unlikely 1 ○ 2 ○ 3 ○ 4 ● 5 ○ 6 ○ 7 ○ Very likely

How confident are you in your answer?

○ Very confident
○ Somewhat confident
○ Somewhat not confident
○ Not at all confident

Figure 3.5: Sample trial progression

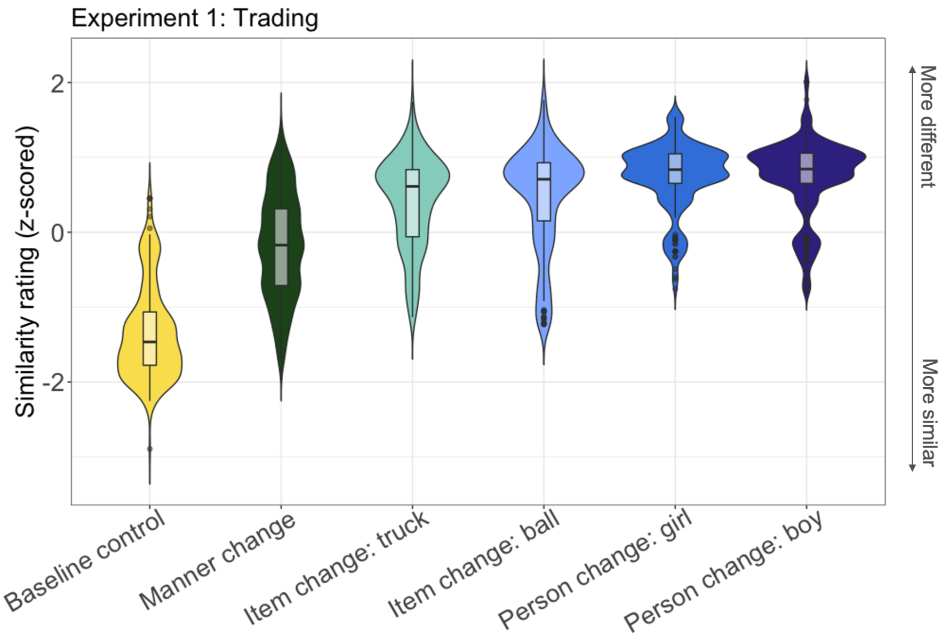


Figure 3.6: Similarity ratings from Experiment 1

3.1.2 Results

Similarity and confidence ratings were z-scored by participant. Any test trials whose similarity rating or time to complete the rating task was more than 2 SD from the mean were excluded (111 outliers excluded for a total of 1473 trials analyzed). The confidence ratings revealed no significant effects and thus will not be discussed further. The z-scored similarity ratings for each condition can be found in Figure 3.6. Lower z-scores indicate a greater degree of similarity, while higher scores indicate a lower degree of similarity.

We used a hierarchical Bayesian analysis on individual trial similarity ratings to model adults' responses by condition. A random intercept for subject was included. We used the average similarity rating (mean = 4.934) and standard deviation (SD = 1.995) of ratings across all conditions as priors for the model. We report the median value of the posterior distribution for each comparison of interest, along with values denoting the upper and lower limits of the 95% credible interval and the probability of direction, from which we can make inferences about the likelihood of the values of the parameter. In our case, a positive effect size corresponds to a larger perceived difference (i.e., one condition is viewed as "more different" than the other). Pairwise comparisons indicated that adults rated all participant changes as more "different" compared to manner changes (β = 1.31, CrI[1.11, 1.51], p(β) = 100% for truck subtraction; β = 1.17, CrI[0.96, 1.37], p(β) = 100% for ball subtraction; β = 1.85, CrI[1.64, 2.06], p(β) = 100% for boy subtraction; β = 1.76, CrI[1.56, 1.96], p(β) = 100% for girl subtraction).

3.1.3 Discussion

We find that changes to all four hypothesized participants of our trading scene were rated as more noteworthy than a change to the manner of motion. This result is predicted under the hypothesis that adults perceived the participant-change conditions as involving not merely a change in the physical event properties, but also a change in conceptual structure: namely, a change between a 4-place and a 3-place concept. All else equal, this suggests that adults perceived our trading scene under a concept with all four participants explicitly represented, as in (12).

Recall from §2.2.1 that adults have a reported visual working memory limit of 3 or 4 (Cowan, 2001; Halberda et al., 2006; Scholl & Pylyshyn, 1999; Sperling, 1960; Trick & Pylyshyn, 1993). Thus, encoding and tracking all 4 participant relations in our trading scene places the visual working memory system at its limit. To alleviate demands on this system, it is possible that people may be representing the scene under two separate event concepts, each with fewer participant relations. This could be interpreted as the implementation of a chunking technique (Chase & Simon, 1973; Feigenson & Carey, 2005; Feigenson & Halberda, 2004; Gobet et al., 2001; Miller, 1956; Moher et al., 2012). For example, rather than being viewed under a single TRADING event concept, our trading scene might be viewed as two sequential GIVINGs :

(16) a. **GIVING1**-AGENT-RECIPIENT-ITEM

b. **GIVING2**-AGENT-RECIPIENT-ITEM

This type of conceptual structure has parallels to previous semantic analyses of TRADE as a two-part predicate composed of a primary GIVING and a symmetrical GIVING (Jackendoff, 1992). We tested this possibility in Experiments 2-4 by comparing the representation of our 'taking' scene to the representation of a different scene – a 'giving-then-disposing' – which is likely to be viewed as two sequential events.

3.2 Experiment 2

We tested perception of another potentially 4-participant event: one actor gives an item to another actor, who then disposes of his or her original item by sliding it to the side (Figure 3.7).

This event type was chosen for two reasons. First, this scene differs minimally from a trading: there are still two actors manipulating two items. As such, it serves as a comparison to determine whether adults are *only* able to track four participants in trading scenes, or if this ability holds

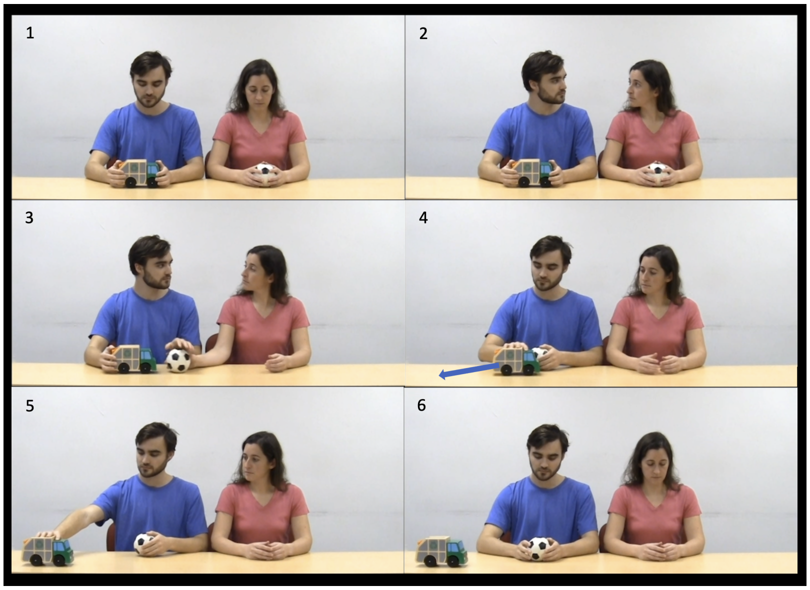


Figure 3.7: Illustration of ‘giving-then-disposing’

across other related scene types. Second, it is plausible that this scene will be viewed under two sequential event concepts – an actor giving an item to the other is followed by the second actor disposing of his or her original item, as in (17). This allows it to serve as a useful comparison to ‘trading,’ to investigate whether a given scene is viewed under one event percept or two.

(17) a. **GIVING**-AGENT-RECIPIENT-ITEM

b. **DISPOSING**-AGENT-ITEM

In Experiment 2, we first ask whether adults track all four participants in our giving-then-disposing scene. This sets the stage for asking whether these participants are perceived in relation to a single event, or in relation to two sequential events. We investigate this second question in Experiments 3-4.

3.2.1 Methods

3.2.1.1 Participants

24 adults (12 female; ages 24-69) were recruited via Prolific. Participants were from the United States or the United Kingdom and were paid $6 for participating.

3.2.1.2 Stimuli

To test the number of participants represented in our ‘giving-then-disposing’ scene, we performed the same manipulation as in Experiment 1, comparing similarity ratings for videos in which a participant was removed and videos in which the manner of motion was changed. The ‘item-removed’ videos consisted of one actor giving an item to the other actor, who took possession of that item before sliding it to the side. The second actor’s original item remained present but unmoved. The ‘manner change’ videos consisted of the actors giving and disposing of their items by lifting them off the table.

In Experiment 1, we saw that the changes to the actors’ participation were noted very strongly by subjects. Thus, we did not include ‘person-removed’ videos in this experiment, under the assumption that the actors would continue to be robustly perceived as event participants. However, to avoid ceiling effects for our crucial manipulations, we included a ceiling control condition: a giving-then-disposing token matched with another token of the same scene which was temporally reversed (i.e., played backwards). Backwards-motion tokens were never paired with their forward counterpart, resulting in 12 total ceiling control trials. Pilot testing confirmed that the temporal reversals were not perceived as unnatural. As in Experiment 1, 12 baseline control trials (consisting of two different giving-then-disposing tokens paired together) were created.

Recording and trial creation was carried out as described in Experiment 1. Each participant saw a total of 60 trials: 36 experimental, 12 ceiling controls, and 12 baseline controls.

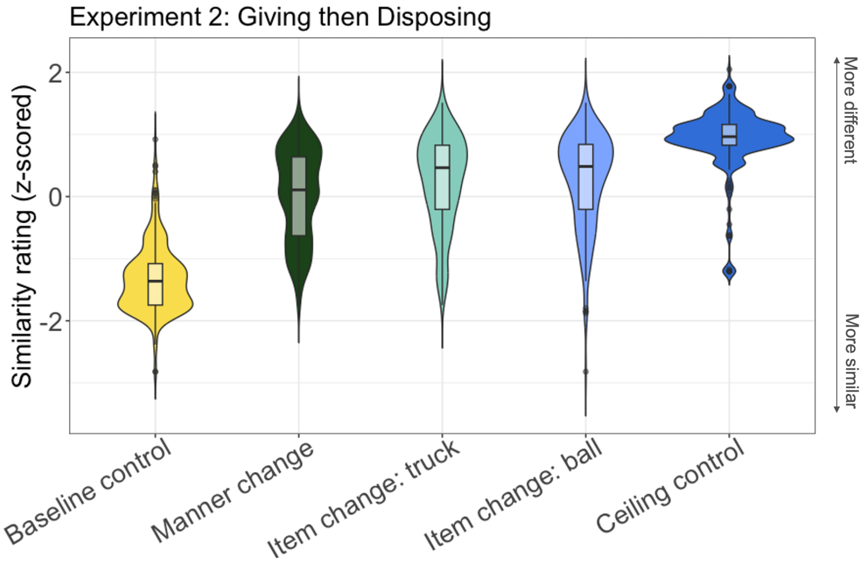


Figure 3.8: Similarity ratings from Experiment 2

3.2.1.3 Procedure

The procedure was identical to Experiment 1.

3.2.2 Results

As in Experiment 1, similarity ratings were *z*-scored by participant; these ratings are shown in Figure 3.8. Also as in Experiment 1, we used a hierarchical Bayesian analysis on individual trial similarity ratings to model adults' responses by condition. We used the average similarity rating (mean = 4.635) and standard deviation (SD = 2.238) of ratings across all conditions as priors for the model. Pairwise comparisons indicated that adults rated all participant changes as more "different" compared to manner changes (β = 0.69, CrI[0.47, 0.92], p(β) = 100% for truck subtraction; β = 0.65, CrI[0.42, 0.88], p(β) = 100% for ball subtraction).

3.2.3 Discussion

The findings in Experiment 2 suggest that adults represent all possible participants in our giving-then-disposing scene. Just as for trading, adults viewed a change in the hypothesized participant structure as more noteworthy than a change in manner of motion, suggesting that the truck and ball filled privileged participant relations in their conceptual representation.

Participants' success in both Experiments 1 and 2 sets us up to test whether they are using a chunking strategy: namely, whether they view the trading scene as two related 'giving' events, and view the giving-then-disposing scene as a 'giving' event that prompts a 'disposing' event. We test this possibility in Experiments 3 and 4 by adopting manipulations from the causal perception literature (Leslie, 1982, 1984). In Experiment 3, we introduce a manipulation to disrupt the sequence of the two hypothesized events, by reversing their relative order. In Experiment 4, we introduce a manipulation to disrupt the coherence of a hypothesized single event percept, by inserting a pause in the middle.

3.3 Experiment 3

In Experiment 3, we manipulated the order in which the potential sub-events occur in our trading and giving-then-disposing scenes. If each scene type is viewed under two sequential event concepts, then changing the relative order of the actions should disrupt this sequence, resulting in a noteworthy difference for the perceiver. But if the scene is viewed under a single event concept, then all else being held equal, changing the order of the actions might be less disruptive.

By hypothesis, we expect giving-then-disposing to be viewed under two event concepts: a GIVING followed by a DISPOSING, as in (17). A change in the relative order of the movements, to a DISPOSING followed by a GIVING, should therefore be a noteworthy difference, due to the reversal of the sequence between the two events. If the trading scene is also viewed under two event concepts – a GIVING followed by a second GIVING, as in (16) – then we would expect a reversal in the order of the two GIVINGs to be similarly noteworthy. This predicts that there should be no interaction of scene type by condition. But if the trading scene is viewed under only one event concept, then we would expect a change to the relative order of movement to be less noteworthy. This predicts an interaction of scene type by condition. Specifically, we would expect order changes to be viewed as less noteworthy compared to manner changes in the trading condition than in the giving-then-disposing condition.

3.3.1 Methods

3.3.1.1 Participants

48 participants (24 female; ages 21-68) were recruited via Prolific. Participants were from the United States and the United Kingdom and were paid $6 for participating.

3.3.1.2 Stimuli

The novel manipulation in Experiment 3 was an order change. For trading scenes, this was achieved by pairing a trading video in which the girl first gives her ball to the boy, with a trading video in which the boy first gives his truck to the girl. In giving-then-disposing scenes, the order change was achieved by pairing the giving-disposing scene described in Experiment 2 with another type of scene developed for this experiment. In this new scene type, one actor first “disposes” of his or her item by sliding it off to the side, after which the second actor “gives” his or her item to the first actor (Figure 3.9).

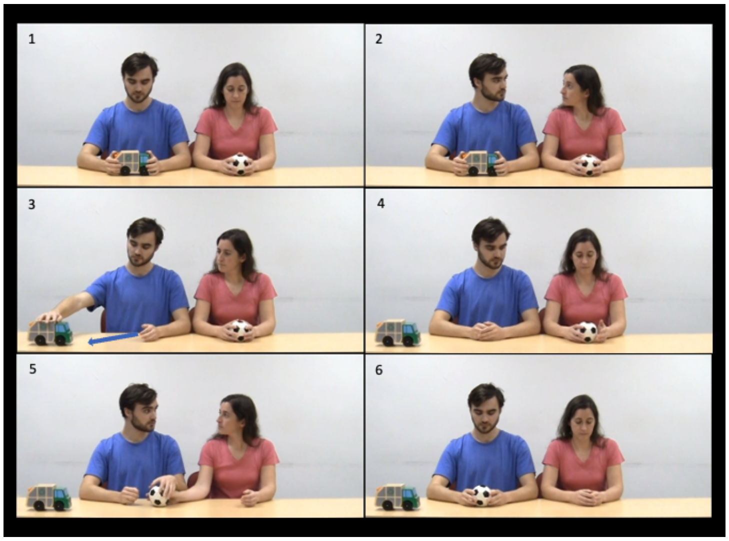


Figure 3.9: Illustration of an “order change” for giving-then-disposing

Baseline control and manner change conditions for each scene type were identical to those in Experiments 1 and 2, respectively. To prevent ceiling effects, a person-removed condition was included for trading scenes (as in Experiment 1), and the backwards-motion condition was included for giving-then-disposing (as in Experiment 2). Although the critical manipulation is the order change, the control and manner change conditions were included to keep the experimental setup as similar as possible to Experiments 1 and 2.

3.3.1.3 Procedure

The procedure was the same as the previous experiments, with the addition of scene type as a between-subjects factor. Half of the participants were assigned to the 'trading' condition and half to the 'giving-then-disposing' condition. The experiment consisted of 60 test trials: 12 baseline controls, 18 manner changes, 18 order changes, and 12 ceiling controls.

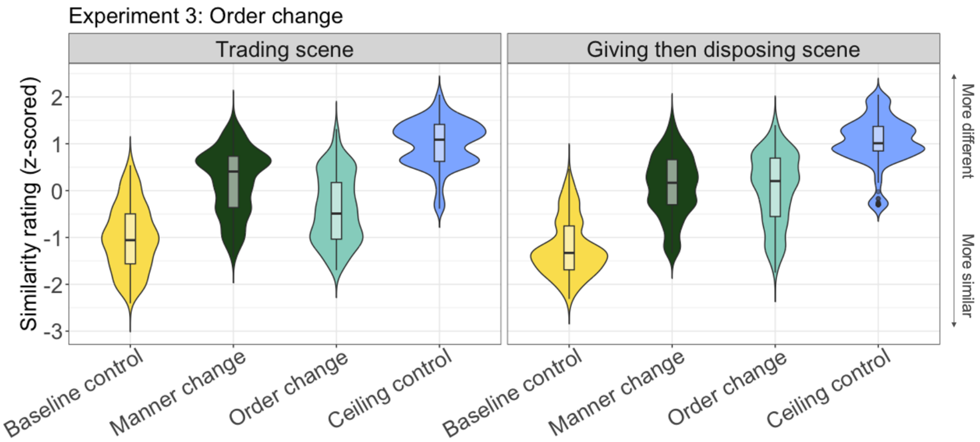


Figure 3.10: Similarity ratings from Experiment 3

3.3.2 Results

The z-scored similarity ratings for Experiment 3 are shown in Figure 3.10. As in Experiment 1, we used a hierarchical Bayesian analysis on individual trial similarity ratings to model adults' responses by condition and scene type and all interactions. We used the overall average similarity rating (mean = 4.47) and standard deviation (SD = 2.18) of ratings across all conditions as priors for the model. A Bayesian model weights comparison revealed that including the condition by scene type interaction term significantly improved model fit, with the model including the interaction receiving virtually all the model weight (1.000) compared to the model without (5.105e-23). Pairwise comparisons indicated that in the trading scene, adults rated the order change as credibly less noteworthy than manner changes (β = -1.48, CrI[-1.68, -1.28], p(β) = 100%). Unlike in trading scenes, order changes were not perceived as less noteworthy than manner changes in giving then disposing scenes: instead, they were viewed as equally noteworthy changes (β = 0.03, CrI[-0.24, 0.17], p(β) = 61.64%).

3.3.3 Discussion

The findings of Experiment 3 suggest that our trading and giving-then-disposing scenes were viewed under different types of event representations. A change to the order of movement was rated as significantly more noteworthy for giving-then-disposing than for trading. For giving-then-disposing, order changes were viewed as just as noteworthy as another physical change to the event (the manner of motion). For trading, order changes were rated as significantly less noteworthy than changes to other event properties. It appears that the giving-then-disposing percept was disrupted by a reversal to the order of motion, but the trading percept was not. This is consistent with the hypothesis that the giving-then-disposing scene is perceived under a two-event structure (a GIVING followed by a DISPOSING), whereas the trading scene is perceived as a single coherent TRADING event, and not as two sequential GIVINGs.

3.4 Experiment 4

Experiment 4 aimed to marshal further support for our interpretation of Experiment 3. Here, we asked whether each event percept would withstand a disruption to its timing. For both the trading and giving-then-disposing scenes, we manipulated the timing with which the possible sequential events occurred by inserting a pause at the hypothesized event boundary, after the first item was moved.

If a scene is being viewed under a single event concept, the insertion of a pause should be viewed as noteworthy. It should break the coherence of the single event percept, causing the scene to be viewed as two sequential events instead. If, on the other hand, the scene is viewed under two event concepts initially, then inserting a pause will not disrupt the event percept as substantially, and thus should not be as noticeable a change. On the hypothesis that our giving-then-disposing scene is viewed as two sequential events, but our trading scene is viewed as one event, then we again predict a condition by scene type interaction. In this case, we expect the interaction to go in the opposite direction as for Experiment 3: the crucial timing manipulation should be viewed as more noteworthy for 'trading' than for 'giving-then-disposing.'

3.4.1 Methods

3.4.1.1 Participants

48 participants (24 female; ages 20-66) were recruited via Prolific. Participants were from the United States or the United Kingdom and were paid $6 for participating.

3.4.1.2 Stimuli

The video stimuli for Experiment 4 were identical to those for Experiment 3, except that we replaced the order change stimuli with 'timing change' stimuli. These timing change stimuli were created by recording new tokens of the trading and giving-then-disposing scenes, now with a 1-second pause after the first actor gives their item to the second actor. As before, an audio recording with cues for each movement alongside a metronome beat was used to ensure the pause was equally long for all tokens and that movement occurred at the same time points throughout the scene. Token pairs were created as in the previous experiments.

3.4.1.3 Procedure

The procedure was identical to Experiment 3.

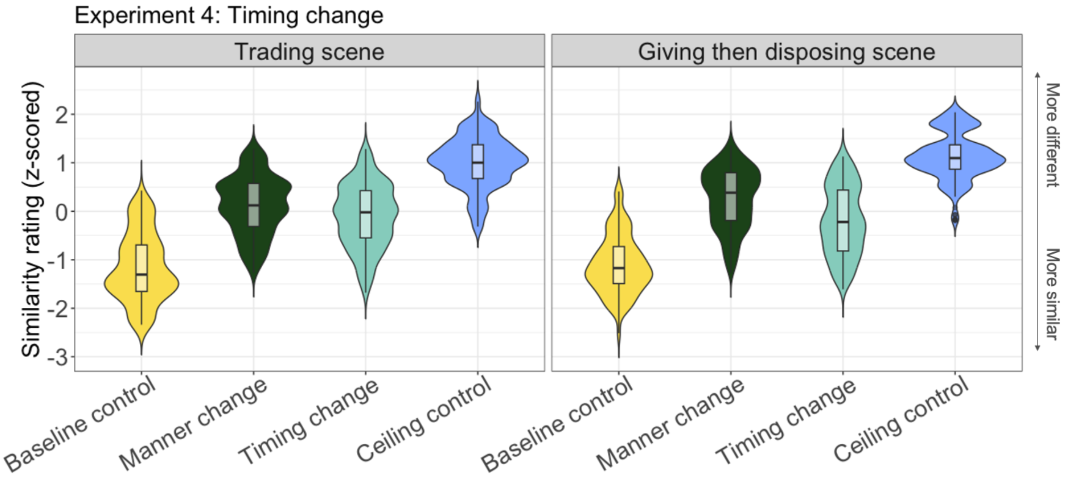


Figure 3.11: Similarity ratings from Experiment 4

3.4.2 Results

The *z*-scored similarity ratings for Experiment 4 are shown in Figure 3.11. As in Experiment 3, we used a hierarchical Bayesian analysis on individual trial similarity ratings to model adults' responses by condition and scene type and all interactions. We used the average similarity rating (mean = 4.69) and standard deviation (SD = 2.23) of ratings across all conditions as priors for the model. A Bayesian model weights comparison revealed that including the condition by scene type interaction term significantly improved model fit, with the model including the interaction receiving virtually all the model weight (1.000) compared to the model without (3.77e-11). As predicted, pairwise comparisons indicated that adults viewed timing changes as more noteworthy in the trading scene than the giving-disposing scene (β = 1.20, CrI[1.00, 1.40], p(β) = 100%). Although timing changes were viewed as less noteworthy than manner changes for both scene types (β = -0.62, CrI[-0.8, 0.42], p(β) = 100% for trading; β = -1.13, CrI[-1.33, -0.94], p(β) = 100% for giving-disposing), timing changes were viewed as more similar to manner changes for trading (median for timing change: 4.95; manner change: 5.57), and less similar to manner changes for giving-then-disposing (median for timing change: 3.75; manner change: 4.88).

3.4.3 Discussion

Consistent with the results of Experiment 3, we again found a predicted condition by scene type interaction, confirming a difference in how our trading and giving-then-disposing scenes are represented. A pause inserted into the trading scenes was seen as more noteworthy than the same length of pause inserted in same position in the giving-then-disposing scenes. This suggests that the giving-then-disposing percept was not disrupted by the insertion of a pause, as predicted under the hypothesis that this scene was already viewed as two sequential events. However, the trading percept was disrupted by the insertion of a pause, suggesting that it was not originally viewed as two separate events.

3.5 General Discussion

This study examined the conceptual representation of a trading scene as a case study of high-adicity event perception. Assuming that these similarity judgements probe event structure in the manner described above, we find that adults view a scene of trading under a 4-participant structure, in which both traders and both traded items are explicitly represented. Moreover, by comparing against another closely related, plausibly four-participant event, we found converging evidence that the trading scene was viewed under one TRADING event concept, rather than as two sequential GIVINGs. These findings are interesting in light of reported constraints on visual perception: in order to represent all four participants in relation to a single event, the visual working memory system may be operating at its reported limit of 4 (Cowan, 2001; Halberda et al., 2006; Scholl & Pylyshyn, 1999; Sperling, 1960; Trick & Pylyshyn, 1993). Our findings suggest that this upper limit is not being circumvented by chunking the trading scene into sequential events, each with fewer participants. Instead, adult visual perception appears capable of yielding a 4-place event concept without this particular type of internal structure.

This finding has potential linguistic implications for semantic analyses of verbs of TRADING that treat this predicate as composed of two sequential GIVINGs (Jackendoff, 1992). Our findings suggest that, in nonlinguistic visual perception, our trading scene was not viewed as two sequential giving events. However, this does not preclude the possibility that this scene may be represented with other types of internal structure. For instance, instead of chunking the scene into two sequential events, people may instead chunk the event participants into groups that bear similar relations to their events: two traders, and two things traded. Alternatively, the scene could be chunked by ownership, with traders grouped with their initial items – see §2.2.2 for a more detailed discussion of these two analyses. These types of grouping may be more in line with other analyses of reciprocal events, in which symmetric predicates are derived from collective meanings (Winter, 2018).

With regards to language acquisition, our findings in this chapter open questions for early verb learning. Acquiring a verb like *trade* requires mapping a linguistic form onto a conceptual representation of an event which falls under the TRADING category. As infants have more stringent caps on visual working memory than adults (Feigenson & Carey, 2003, 2005; Feigenson et al., 2002), does the perceptual system of a young verb learner likewise readily yield 4-place representations of trading scenes? If so, we might ask what mechanisms they deploy to circumvent their visual working memory limits. If not, this would raise a puzzle for how verbs like *trade* are acquired. In the next chapter, we investigate whether the perceptual system of young children yields a similar 4-place representation of trading scenes.

CHAPTER 4

Perception of complex event types in preschool-aged children

4.1 Overview

In this chapter, we report findings from a study with preschool-aged children adapted from Experiment 1 of the adult series (§3.1), as an important first step towards understanding the developmental trajectory of visual working memory as it interacts with event representation for high-adicity concepts like TRADE. To identify young children's conceptual representations of a trading scene, we conducted a "picky puppet" study (Waxman & Gelman, 1986) with children between the ages of 3.5 and 5.5, using the same stimuli as in Experiment 1 of the adult series. We find that children rated the participant changes as "different" significantly more than the corresponding manner changes, suggesting that preschoolers can track all 4 participants in a trading scene. We discuss the implications of these findings as they relate to early verb learning with high-adicity concepts, and in particular, to acquiring the meaning of the verb *trade*.

4.2 Methods

Children were presented with pairs of videos across the same four conditions as in Experiment 1 in Chapter 3. All conditions consisted of a trade token paired with another video token. For the control condition, the trade was paired with another token of the same trading event type, in which the toys were exchanged by sliding them across the table. In the manner change condition, it was paired with an arcing trade, in which the actors lifted the toys in an arc to exchange them. In the item change condition, a trade token was paired with a token in which one of the items was no longer moved – the same item was passed back and forth between the two actors. Lastly, the person change condition consisted of a trade token paired with a video in which one of the actors was no longer involved – one of the actors looked off to the side while the other swapped the two toys.

Children were trained to judge whether a picky puppet would ‘like’ a particular video pair based on the fact that the puppet only likes when the same thing happens in the two videos. Children were then prompted to give their judgements for each video pair. By allowing the children to judge the videos, this provides a measure of whether they consider the videos to be similar or different. If children are able to track all four participants in the scene, they should rank both types of participant changes (item and person) as “different” more frequently than the manner change. If they are not able to track all four participants, we would expect to see no such asymmetry in “different” rankings for at least one of the participant changes compared to the manner change.

4.2.1 Participants

Of the 48 children tested and included in the analysis, 22 (11 female) were recruited through and tested at a preschool in Los Angeles. The remaining 26 children (15 female) were recruited online or over the phone through the University of California, Los Angeles (UCLA) Developmental Subject Pool. Informed parental consent was obtained in accordance with the protocols of the UCLA Institutional Review Board. All children were tested individually, either at their school or in the lab, in a quiet area (conference or testing room).

The children ranged from 3.5 to 5.5 years old (range: 3;5;15 - 5;4;21, mean: 4;6;6). An additional 20 children were tested but excluded from the analysis due to failure on the practice trials (9 children), sticker/response bias (giving the same response on all but one of test trials; 5 children), missing both trials in a condition (1), or missing 4 or more trials (5). “Missing” a trial was defined as the child being distracted for one or both of the videos in a pair, meaning that they did not see one of the tokens they were meant to compare.

4.2.2 Stimuli

The video stimuli were the same as in Experiment 1 of the adult study, described in Chapter 3. In a given trial, two video tokens were paired together with 1s of black screen between them. As in Experiment 1, test stimuli consisted of controls, manner changes, item changes, and person changes. The stimuli were presented using Microsoft PowerPoint. Videos played automatically and were followed by a black screen, during which the child was prompted for a response. Only videos presented during the training and practice phases could be replayed if the child asked.

Participants were tested in a mixed within- and between-subject design. Children saw 10 test trials, of which only the last 8 were analyzed (described in more detail below). This number of trials was determined by pilot testing to be appropriate for children of this age. The order of presentation was counterbalanced within-subjects such that one video pair for a given condition began with a trade token and the other began with the change. Half of the children saw only video pairs where the truck moved first, while the other half only saw video pairs where the ball moved first. Unlike in the adult experiment, each child saw only one type of person change (i.e., boy or girl removed) and one type of item change (i.e., truck or ball removed) during the test phase. The between-subject subconditions were each arranged into 4 pseudo-randomized trial lists, and children were randomly assigned to one of four trial orders for a randomly assigned subcondition.

4.2.3 Procedure

Each child was brought to a quiet testing location and asked to sit at a table with a laptop computer, a grid for placing stickers, a sheet of happy and sad face stickers, and a smaller sheet of star stickers with another small grid. At the start of the experiment, the experimenter introduced the children to the ‘activity’, beginning by introducing a puppet named Miss Hippo. The children were told that Miss Hippo is a ‘very picky hippo’, who only likes things that are the same and really dislikes things that are not. The experimenter then explained that the happy stickers will be placed on the grid for videos that have ‘the same thing happen’, and sad stickers will be placed on the grid for videos that have ‘something different happen’ (for a full version of the script, see Appendix A). The children then saw 4 training trials, where they watched 2 “same” video pairs and 2 “different” video pairs. The training videos consisted of a girl interacting with a bottle of orange liquid, either shaking, tapping, or spinning the bottle. “Same” video pairs were those in which two different tokens of the same action were displayed, while “different” video pairs were those in which tokens of two different actions were displayed. After watching each practice video pair, Miss Hippo gave an emphatic judgement, cheering for ‘same’ pairs and expressing disgust for ‘different’ pairs. The children were also prompted to place the corresponding happy or sad sticker on the grid based on Miss Hippo’s reaction.

Participants then moved into the practice phase. The typical practice phase consisted of 2 video pairs, with one match and one mismatched pair. These videos depicted a girl interacting with a box, either opening it or flipping it on its side. Miss Hippo told the children that she is very sleepy (due to missing her nap). She then “fell asleep” and the experimenter placed her on the table next to the other materials. The experimenter then prompted the children to watch the next video pair to see if the same thing happens before Miss Hippo “wakes up”. The children were prompted to make a judgement of “same” or “different”. If the children did not answer or said that they did not know, the experimenter replayed the video pair for them. Once the children gave a judgement, the experimenter “woke up” Miss Hippo, who then asked to watch the videos again and gave her own judgement. If the children got both of the judgements correct, they moved into the test phase. If the children got one or more judgements incorrect, Miss Hippo would explain why she gave her judgement, and then proceed into a second practice phase, consisting of 2 more video pairs with one match and one mismatch. The same procedure was followed for these additional practice phases. If the children gave correct judgements for both of these pairs, they moved to the test phase and were included in the analysis unless they met other exclusion criteria. If, instead, the children gave an incorrect judgement for one or both of these pairs, they proceeded to a shortened version of the test phase and were not included in the analysis.

The test phase began with Miss Hippo asking if she could take her nap and if the children could watch the rest of the videos without her. The experimenter then placed the puppet to the side as before, explaining that Miss Hippo is "very tired", and reminding the children of the rules for placing the happy and sad stickers. The experimenter then played the first video pair, consisting of two lifting-trade tokens (i.e., the same tokens used to create manner change videos) paired together, and asked the children for a judgement. After helping the children place the corresponding sticker on the grid, the experimenter praised the children and played the next video pair, consisting of a lifting-trade and control token paired together. The experimenter's reaction was the same (e.g., praise) regardless of the judgement that the children provided. These first two trials of the test phase were not analyzed, but included in the procedure so the children could adjust to visual differences between test stimuli and training stimuli (e.g., containing two actors compared to one actor, different location, inclusion of different objects, etc.).

Phase	*# of trials*	*Response from*	*Feedback*
Training	4	Miss Hippo	N/A
Practice	2 or 4	Child	Yes
Test	10	Child	No

Table 4.1: Experiment design and trial sequence

This procedure continued until all of the remaining 8 test trials had been played. After the conclusion of the test phase, the experimenter "woke up" Miss Hippo, who then praised and thanked the children for their help. The training phase consisted of 4 trials, the practice phase of 2 trials (or 4, if necessary), and the test phase of 10 trials, for a total of 16 (or 18) trials. A summary of the trial structure for the experiment is shown in Table 4.1. To ensure continued attention throughout the experiment, after every 4 video pairs a spinning gold star appeared on the screen and the children were prompted to place a star sticker on the smaller grid.

4.3 Results

4.3.1 Response coding and trial exclusions

When the children were prompted for a response during the test phase, their first response was recorded whether it was verbal (the child saying "same" or "different") or an action (reaching for the corresponding "same" or "different" sticker). If the children were silent or did not respond, they were prompted to choose which sticker they thought fit the videos. Occasionally, children's verbal responses and actions were not consistent or they changed their minds before committing to a sticker. For this reason, both initial and final responses were recorded. However, this response change was very rare (only 9 trials out of the 351 included for analysis). Analyzing the initial responses did not alter the overall pattern of results compared to the final responses. For this reason, only final responses are reported in the next section. Additionally, any trials in which a child failed to attend to one or both video tokens in a pair were excluded from the analysis (33 trials were excluded for this reason, yielding a total of 351 trials analyzed).

4.3.2 Response analysis

The averaged ratings for each condition can be found in Figure 4.1. Averaged ratings were computed by calculating the mean response for a child for each condition (with 1 being "different" and 0 being "same"), then averaging the means across participants. We plot the averaged ratings for visualization purposes only; analyses were conducted with individual trial responses as the dependent measure.

We used a hierarchical Bayesian analysis on individual trial responses to model children's responses by condition. A random intercept for subject was included. We used the average response (mean = 0.5005) and standard deviation (sd = 0.48718) of responses across all conditions as priors for the model. We report the median value of the posterior distribution for each comparison of interest, along with values denoting the upper and lower limits of the 95% credible interval and the

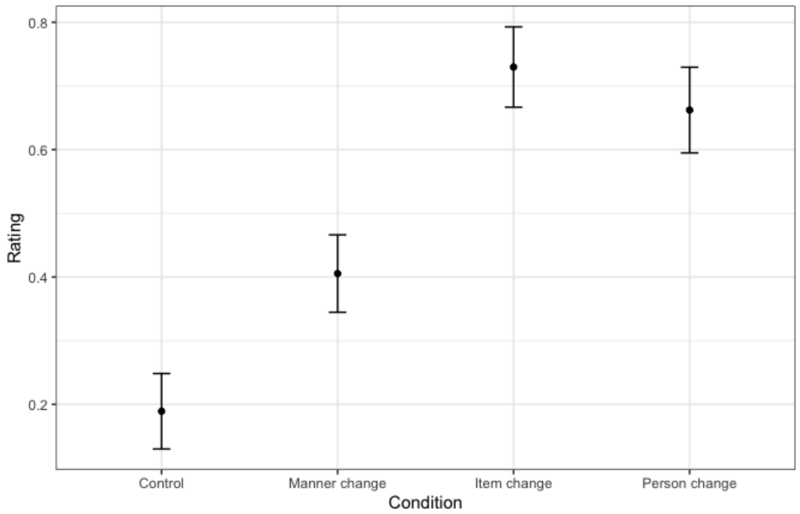


Figure 4.1: Averaged responses for participants. Error bars represent standard errors of the mean.

probability of direction, from which we can make inferences about the likelihood of the values of the parameter. In our case, a positive effect size corresponds to a larger perceived difference (i.e., one condition is viewed as "more different" than the other). An aggregate effect size is credible if the interval does not include zero. We also report the posterior probability of the effect size, which corresponds to the proportion of credible values above zero and represents the probability of having any nonzero effect. Pairwise comparisons indicated that children rated the truck, ball, and girl subtractions as credibly more "different" compared to manner changes (β = 0.27, CrI[0.09, 0.44], $p(\beta)$ = 0.9983 for truck subtraction; β = 0.19, CrI[0.03, 0.36], $p(\beta)$ = 0.9886 for ball subtraction; β = 0.22, CrI[0.06, 0.39], $p(\beta)$ = 0.9962 for girl subtraction) and a near-credible difference for the boy subtraction (β = 0.14, CrI[-0.03, 0.31], $p(\beta)$ = 0.9430).

Although including an interaction with age did not improve model fit, we show the average responses broken into two groups of younger and older children in Figure 4.2 for completeness.

This visualization allows us to see that both groups of children show the same numerical trend for the relevant changes (with participant changes viewed as more "different" than manner changes), but unlike older children, younger children appear to be at chance for the control condition.

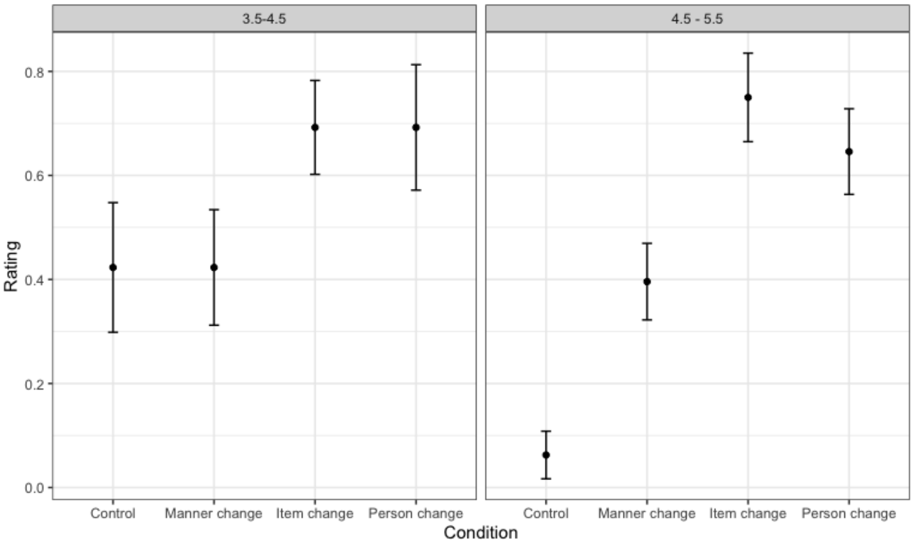


Figure 4.2: Averaged responses broken down by age, with average responses for 3.5-4.5 year-olds on the left and 4.5-5.5 year-olds on the right. Error bars represent standard errors of the mean.

4.4 Discussion

The findings in this experiment point to young children's ability to represent all 4 participants in a trading scene. In particular, the fact that children rate both types of item changes (ball removed and truck removed) and both types of person changes (girl removed and boy removed) as credibly or near-credibly more different than manner changes suggests that these four participants are explicitly encoded within their event representation for this scene. This finding indicates that children behave the same way that adults do on an age-appropriate version of the task in Chapter 3, suggesting that they may represent this scene in the same way that adults do. In other words, these results suggest that children's visual working memory system is capable of yielding a 4-place event concept at 3.5 to 5.5 years old.

The ability for the visual working memory system to yield a 4-place event percept at this age is important for our theories of verb learning, since understanding children's scene concepts is a necessary first step to understanding how they bootstrap high-adicity verb meanings. Namely, prior to drawing conclusions about mechanisms by which children may be mapping between conceptual and linguistic structure on the basis of behavioral evidence, it is important to ensure that children are viewing a stimulus scene the same way adults do, and under the same representation. Recall from the discussion in §2.1.3 that children familiarized with intransitive sentences looked equally to causal and non-causal scenes at test. This was in contrast to cases where children were familiarized with transitive sentences, which led the children to preferentially look at the causal scene. The difficulty interpreting children's ambivalence when presented with intransitive sentences is in part due to the indeterminacy of children's scene representations in earlier bootstrapping work, as their responses could be the result of children viewing the scene under a different representation than the experimenters had presumed. After all, any given event can be viewed under many different concepts, as well as labelled in many different ways. Thus, the experimental investigation of bootstrapping mechanisms requires an understanding of children's representations of a particular stimulus scene prior to drawing conclusions about their behavioral responses. In this experiment, we set the stage for understanding how children can map language to this scene concept for the purpose of bootstrapping verb meanings. Here, we demonstrated that young children represent all four participants in a trading scene, just as the adults did in Experiment 1.

There is some evidence that this task may have been difficult for the younger children: namely, children in the younger cohort were almost equally likely to judge the controls "different" as they were to judge them as the "same" (mean = 0.42, SD = 0.45). Although this task is designed for preschool children, a good deal of reasoning and recall is nonetheless required to make a judgement. The child needs to not only maintain the events in memory in order to determine whether any changes occur, but also to remember that the puppet likes only those pairs that have the same thing happen. Additionally, children must remember that the happy stickers correspond to "same" and that sad stickers correspond to "different". Given this lengthy reasoning chain, it is possible that younger children simply have difficulty with the task as a whole, despite being able to maintain the participants in visual working memory. For this reason, an easier task with more implicit measurements may be a better indicator of young children's ability to represent all four participants in a trading scene. Ongoing work is currently testing infants with such an implicit measure, using the same habituation-based design as in Perkins et al. (2024) with the trading scenes used in this experiment.

The overall findings of this study suggest that children between 3.5 and 5.5 years old are able to represent all 4 participants in a trading scene. However, it is still unknown whether they do so by representing the trading scene under a single TRADE concept (18a), or whether they deploy some form of chunking to avoid placing the visual working memory system at (or beyond) its limit. In the experiments from Chapter 3 we found that adults do not appear to chunk the same trading scene into two sequential GIVINGS (18b), but it remains a possibility that children could be chunking in this manner. Similarly, it is an open question whether children (or adults) deploy alternative chunking strategies, such as chunking by participant type (18c) or by ownership (18d):

(18) a. **TRADE**-AGENT1-AGENT2-ITEM1-ITEM2 Not chunked

b. **GIVING1**-AGENT1-AGENT2-ITEM1 Chunked into sub-events
GIVING2-AGENT1-AGENT2-ITEM2

c. **TRADE**-[AGENT1-AGENT2]-[ITEM1-ITEM2] Chunked by participant type

d. **TRADE**-[AGENT1-ITEM1]-[AGENT2-ITEM2] Chunked by ownership

Further work will be needed to adjudicate between these chunking possibilities for children. Deploying one of these chunking strategies could have implications for learning the meaning of a verb like *trade*, which we will now consider in more detail.

4.4.1 Implications for bootstrapping

A sentence describing a scene may refer to many different aspects of that scene. Given this indeterminacy, bootstrapping could provide a useful heuristic to narrow down possible scene referents. For example, children may see a variety of different ‘trading’ scenes and could hear *trade* in a variety of different sentences. Assuming that at least some of these ‘trading’ scenes can be perceived with 4 participants (as shown in the current study), it is worthwhile to ask what information is present in the syntactic distributions in which children hear verbs like *trade*. Does it always occur in a 4-place frame, or is the actual distribution more varied?

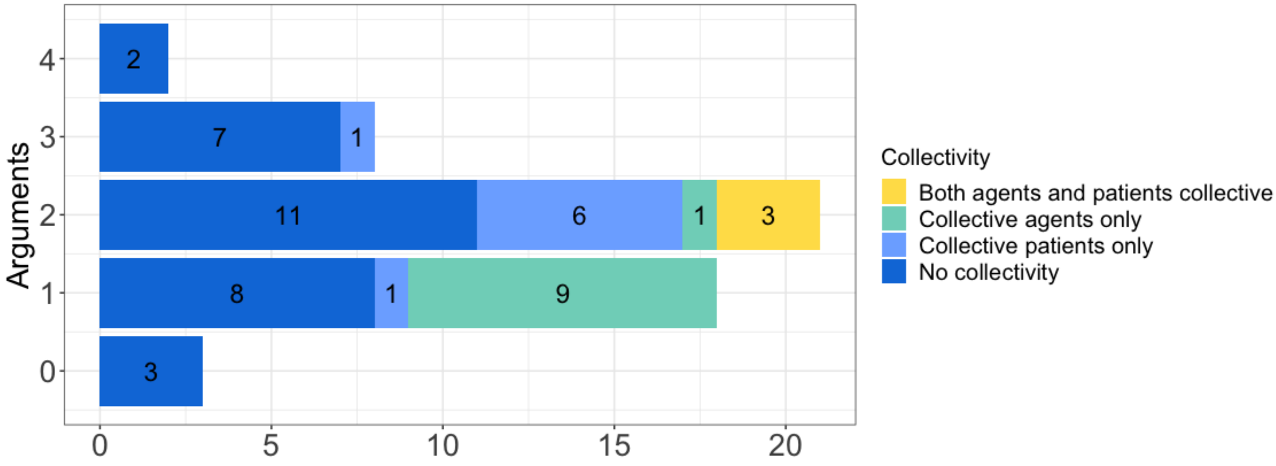


Figure 4.3: Distribution of *trade* in North-American English child-directed speech. Utterances with 0 arguments consisted of standalone uses of *"trade"* or those with only implicit arguments.

A survey of the North American English corpora in the CHILDES TalkBank (MacWhinney, 2000), filtered for speech directed at typically-developing children, yielded only 52 utterances containing *trade* out of 637,818 utterances spoken to or around the target child (c.f., 6,368 instances of *take* and 4,337 of *give*). The frequencies of different uses of *trade* within the corpus can be found in Figure 4.3.

Of the 52 utterances, only 2 existed in a full 4-place frame, as in (19) and (20) where both agents and both patients are explicitly named.

(19) FAT: I'll trade you my Brussels sprouts for my chicken (Demetras 1989, 020612.cha)

(20) MOT: I'll trade you a hole one for a solid one (Post 1994, 020503.cha)

Another 8 utterances occurred in ditransitive frames (21), of which 5 labelled both agents but only one patient as in (21a), and 2 labelled both patients but only one agent, as in (21b):

(21) a. LOI: I'll trade you cookie (Bloom 1970, 010807.cha)

b. MOT: and Jack trades the pig for a bean (Brown 1973, 030826.cha)

c. SIB: Kalie I'll trade ya babies (Post 1994, 020503.cha)

One ditransitive utterance, (21c) refers to collective patients (in the context of trading two baby dolls). Overall, sentences with collective patients or agents were frequent, constituting over 40% of the utterances with *trade.* Of these, 3 contained both collective agents and patients, 8 more contained collective agents, and 10 referred to collective patients. The frequencies of these collective forms can be found in Table 4.2.

Collective	*Number of utterances*
Both agent and items	3
Agents only	8
Items only	10
None	31

Table 4.2: Frequency of collective agents and patients in utterances containing *trade*

The majority of utterances containing *trade* occurred in transitive 2-argument frames (23 utterances), as in (22), or with only a single argument (22 utterances), as in (23):

(22) a. MOT: I'll trade you (Brown 1973, 020100b.cha)

b. MOT: do you wanna trade Mommies (Suppes 1979, 030107.cha)

c. INV: well they traded places (Clark 1979, 021002b.cha)

(23) a. MOT: trade me (Brent 2001, 010007.cha)

b. MOT: we'll trade (Sachs 1983, 020508.cha)

c. MOT: she might not want to trade (Post 1994, 020024.cha)

It is evident that *trade* does not occur frequently in child-directed speech, and even less frequently does it occur in its full 4-place frame. The prevalence of transitive and intransitive uses of *trade* bears upon questions of how children can bootstrap themselves into the meaning of this complex verb. The findings of our experiment suggest that children encode all four participants in at least some scenes of trading. One bootstrapping hypothesis proposes that children expect arguments and participants to match one-to-one (Fisher, 1996; Lidz & Gleitman, 2004; Naigles, 1990; Yuan et al., 2012, among others). If this is the case, they may expect that all of the perceived participants in an event should also be realized as clause arguments (see the discussion of 'one-to-one matching' in §2.1.3). However, under this hypothesis, the infrequency of 4-place uses of *trade* in the input would pose significant challenges to young learners as they attempt to map between sentences with this verb and a 4-place event concept. If, on the other hand, they represent the trading scene as two sequential GIVINGs, this may not pose as much of a challenge, given that *trade* occurs in 3-argument frames more often (though still not as frequently as it appears in 1- and 2-argument frames). Thus, although the input would still not be fully conducive to this type of one-to-one matching, the 3-argument frames would be encountered more frequently and thus alleviate some of the challenge for the learner. However, as discussed in §2.1.3, Perkins et al. (2024) argue that children do not expect one-to-one alignment between arguments and participants, and instead more flexibly link scene and sentence percepts based on thematic content. If children are relying on this type of 'thematic linking', this could mean that the lack of 4-place uses of *trade* in children's input would not pose a significant challenge in and of itself, provided children can link the syntax they are hearing with their scene representation in other ways.

The ways that children may be able to link the syntactic distribution of *trade* with their representation of TRADING scenes may depend on whether they perceive these scenes with internal structure along the lines of those in (18), repeated below:

(18) a. **TRADE**-AGENT1-AGENT2-ITEM1-ITEM2 Not chunked

b. **GIVING1**-AGENT1-AGENT2-ITEM1 Chunked into sub-events
GIVING2-AGENT1-AGENT2-ITEM2

c. **TRADE**-[AGENT1-AGENT2]-[ITEM1-ITEM2] Chunked by participant type

d. **TRADE**-[AGENT1-ITEM1]-[AGENT2-ITEM2] Chunked by ownership

In other words, a particular kind of syntactic frame could link to a particular kind of chunked representation. Thematic linking relies upon arguments aligning with participant relations in the right way – namely, the subject labels the perceived agent, the object labels the perceived patient, and so on. Under thematic linking, if children view the trading scene as two sequential GIVINGS as in (18b) and then hear *trade* used in many sentences with 3 or fewer arguments, they may think that the verb labels one of the two GIVING events that they readily perceive in the scene. This could lead children to erroneously conclude that *trade* means *give*, only to realize the error upon hearing the much rarer 4-place frame and noticing that this frame has more arguments than the three they perceive in a GIVING.

Alternatively, if children pattern like adults and do not chunk the scene into two sequential GIVINGs, using one of the other chunking strategies instead, they could be led to other conclusions about the meaning of *trade*. Namely, since children hear sentences with collective uses of *trade* relatively frequently, these sentences could link naturally to a scene percept in which either the two traders and two items are chunked (18c) or the traders and their initial items are chunked by ownership (18d), as described in §2.2.2. For example, sentences like *'I'll trade ya babies'* (21c) could lend themselves to this type of mapping if children realize that this collective syntax picks out a 4-place TRADE concept that they had readily chunked along the lines of (18c) or (18d). Namely, because reciprocals necessarily describe a symmetric event, this could potentially be used as a linking principle to allow children to learn that *trade* refers to TRADING, with children specifically "zooming in" on a construal of the scene in which two agents stand in the same thematic relation to the event, and also two items stand in the same thematic relation to the event. This could lead them to realize that despite appearing very rarely in a 4-place frame, *trade* is really a 4-place predicate that may have an alternation between a reciprocal collective and symmetric 4-place construction. Thus, finding evidence that children perceive the trading scene in this experiment as symmetric could provide evidence that children may be capable of deploying this type of alternation-symmetry linking. This type of linking principle may extend to other collective and symmetric predicates, including *share NP (with)*, *date*, *match*, *marry*, *sibling*, among others (for a full set of such predicates, see Winter, 2018 and Gleitman et al., 1996): because symmetrical and reciprocal predicates generally pattern together, noticing a reciprocal alternation could allow children to recognize that the event being described belongs to this privileged symmetric event class.

Further work exploring the internal organization of children's scene percepts may help identify the mechanisms by which children acquire verbs describing reciprocal concepts like *trade*. This type of work probing the non-linguistic representations of trading scenes at the ages when infants are most actively learning verbs – and when the visual working memory system is known to have a limit of 3 – will be particularly crucial for our understanding of this type of acquisition, and is currently part of our ongoing work.

4.5 Conclusion

In this chapter, we reported findings from an age-appropriate version of Experiment 1 in the adult series. We found that, just like adults, young children between the ages of 3.5 and 5.5 are able to encode all four participants in the same trading scene. This indicates that children's visual working memory systems are capable of yielding a 4-participant event concept at this age. Because learning verbs requires mapping between conceptual representations and linguistic structures, identifying the non-linguisitic conceptual representation is crucial for our understanding of how children may come to acquire verbs like *trade*. Namely, identifying the conceptual representations under which children view scenes like the trading scene in this experiment can inform us about the types of bootstrapping mechanisms that could be deployed to learn these verbs.

Our survey of child-directed speech highlighted the rarity of 4-place uses of *trade* in a child's typical input. If children indeed represent at least some trading scenes under a four-participant concept, a bootstrapping mechanism that relies on one-to-one number matching of arguments to participant roles may pose challenges for the young learner attempting to learn this verb. On the other hand, if (unlike adults) children represent at least some scenes of trading as two related GIVINGs, this may not pose as much of a challenge, given that *trade* occurs in sentences with 3 arguments more frequently (Figure 4.3). However, Perkins et al. (2024) provides evidence that toddlers rely on thematic linking, and not one-to-one matching, to identify a novel verb's referent. Thus, if children instead rely on links between thematic roles and participant relations that they perceive in an event to acquire the verb, as per Perkins et al. (2024), different conceptual structures may map to more or less frequent sentence frames in the input. Understanding how children represent high-adicity events like trades can reveal additional types of linking mechanisms by which children can bootstrap themselves into verb meaning. For example, if children realize that a reciprocal *trade* exists in an alternation with the binary-collective form, this could cue the realization that *trade* describes a symmetric event. Further adjudication between the bootstrapping mechanisms that children use to acquire *trade* (and other verbs that label high-adicity concepts) therefore relies on understanding the types of scene percepts yielded by a child's developing visual working memory system. To our knowledge, this study is one of the first to investigate children's representations of high-adicity event types like TRADE, and serves as an important first step towards more nuanced investigation of verb-learning strategies as they apply to verbs that label complex event types.

Part 2

CHAPTER 5

Sentence processing across development

5.1 Introduction

In Part 1 of this dissertation, we investigated how extralinguistic perceptual systems could potentially limit verb argument structure acquisition through the interaction of the visual working memory system and event perception and representation. Once children have acquired the argument structure for a particular verb, the question arises regarding how they process sentences that use that verb. This speaks to broader questions about whether and how children make use of linguistic knowledge immediately upon acquiring it (see Lidz, 2023). In Part 2 of this dissertation, we examine the somewhat surprising fact that relying on knowledge of a verb's argument structure to guide parsing predictions may actually interfere with the ability to use bottom-up information during sentence processing. This interference appears to stem from interactions of limited working memory and cognitive control systems. Consequently, we will now turn to the question of how broader extralinguistic cognitive systems interact with existing knowledge of verb argument structure in early sentence processing. This chapter provides an overview of sentence processing in both adults and children, as well as a broad overview of how two cognitive systems, inhibitory control and working memory, interact with sentence processing both in adults and across development.

5.2 Sentence processing in adults and children

5.2.1 Characteristics of adult sentence processing

The psycholinguistics literature has broadly converged on several key facts about the adult parser. In particular, adult sentence processing appears to be incremental and predictive, meaning the parser does not wait for all words to be heard prior to postulating a meaningful interpretation, and forms hypotheses about upcoming words and phrases that have not yet been encountered (Bever, 1970; Duffy et al., 1988; Frazier, 1979; Kimball, 1973; Simpson, 1984; Tanenhaus et al., 1979; Trueswell & Tanenhaus, 1994, among others). In this section, we will review some key pieces of evidence that support this.

A significant portion of the language processing literature has concerned so-called "garden-path sentences": sentences with temporary structural ambiguities that lead readers (or listeners) 'down the garden path' to an incorrect interpretation that must then be revised. These sentences have highlighted an important aspect of how adult language speakers and listeners use their linguistic knowledge over the course of processing a sentence: the processor incrementally builds structure while incorporating each word, leading to a specific interpretation. Occasionally, these interpretations are incorrect, leading to the "garden-path effect". For example, in the reduced relative clause sentence (24):

(24) The horse raced past the barn fell. (Bever 1970)

adult readers initially interpret *raced past the barn* as the main verb phrase. However, the correct interpretation of this sentence is one where *raced past the barn* is a reduced relative clause, modifying *the horse*. The fact that the sentence is parsed incrementally leads to an error once the comprehender encounters the word *fell*, as there is no grammatical structure that could be constructed from the currently considered structure that integrates this word.

Evidence for this type of processing difficulty comes from a wealth of studies investigating the online processing of such "garden-path" sentences. For example, in an eye-tracking study of adult readers, Frazier and Rayner (1982) presented participants with sentences like (25):

(25) Since Jay always jogs a mile seems like a short distance to him.

If readers use an incremental parsing strategy, incorporating each word individually as it is read, they would initially interpret *a mile* to be the direct object of the verb *jogs*. However, this analysis renders the following word, *seems*, incompatible with the currently considered structure. Instead, to correctly parse this sentence, readers must reinterpret *a mile* as the NP subject of the next clause. Frazier and Rayner (1982) found that readers spent longer periods of time fixating on the disambiguation point (*seems*) than for minimally different control sentences like *Since Jay always jogs a mile* ***this*** *seems like a short distance to him.* This delay is not predicted under alternative parsing strategies, such as if the parser waited until the end of the sentence to commit to a structure. If this were the case, no slow-downs in the disambiguation region would be observed: the parser would not have committed to any structure and thus would not need to revise its commitments. Adults' online processing performance with garden-path sentences like (24) and (25) has been replicated in a wealth of studies (Altmann, 1986; Rayner et al., 1983; Rayner et al., 1992, among others), which has led to the conclusion that sentence processing occurs incrementally.

Sentence processing relies heavily on existing linguistic knowledge and appears to be influenced by lexical and semantic information. For example, in sentences containing an ambiguously-attached PP, like (26):

(26) Anne hit the thief with the stick.

there are two possible interpretations of the prepositional phrase: in one interpretation, the stick is a modifier of *the thief* and is attached to the second NP ("NP attachment"). In the second interpretation, the stick is an instrument, and is consequently attached to the VP (known as "VP attachment"). Adult comprehenders have a strong preference for the VP attachment interpretation when encountering the ambiguous preposition *with* in globally ambiguous sentences like these (e.g., Rayner et al., 1983; Taraban & McClelland, 1988). Altering the last noun in the sentence, as in (27), leads to parsing difficulties:

(27) Anne hit the thief with the wart.

Here, the PP can only be NP-attached (after all, it is not possible to use a wart as an instrument of hitting), violating the initial preference for VP attachment. The difficulties in parsing a sentence like (27) may arise due to the selectional properties of *hit*, which often includes an Instrument introduced by the preposition *with* (Taraban & McClelland, 1988). To determine whether PP-attachment preferences could be influenced by lexical information, Spivey-Knowlton and Sedivy (1995) used a self-paced reading task to test adults' performance on V-NP-*with*-NP sentences like (26) and (27) with different verbs. They found that for sentences with action verbs like *blew open*, adults had a strong VP-attachment preference, but for sentences with psychological predicates or verbs of perception like *look at*, participants instead have a preference for NP-attachment. These findings suggest that adults use lexical information to guide their parsing.

Similarly, real-world knowledge and visual or pragmatic context can facilitate processing, even with sentences that are difficult to parse in isolation (such as garden-path sentences). For example, Altmann and Steedman (1988) provided participants with additional context prior to presenting them with either NP-attached or VP-attached target sentences. The contexts supported NP-attachment (e.g., two possible NP referents mentioned in the context preceding) or VP-attachment (e.g., only one relevant NP mentioned in the context preceding). They found that the preceding context influenced the attachment preferences of readers – if they had read the NP-supporting context, their reading time for VP-attached target sentences increased compared to NP-attached target sentences. This suggests that adults incorporate contextual information as they are incrementally processing sentences, using context to guide their parsing.

Adults are also able to use visual context to guide parsing. Tanenhaus et al. (1995) equipped adults with head-mounted eye-trackers and presented them with visual layouts with four quadrants.

In one condition, the visual layout contained two possible referents (for example, two apples), one goal destination and one distractor. In the other, the visual layout contained only one possible referent (for example, an apple and a pencil) as well as the same goal destination and a distractor. The adults were then asked to follow instructions that were either temporarily ambiguous, as in (28a), or unambiguous (28b). Here, (28a) is temporarily ambiguous because the PP *on the towel* could initially be interpreted as either being VP-attached (the destination for the apple) or NP-attached (a modifier of the apple).

(28) a. Put the apple on the towel in the box.

b. Put the apple that is on the towel in the box.

In the one-referent condition, adults presented with ambiguous sentences looked to the distractor location (another towel), suggesting they were incrementally processing the sentence, committing to an analysis in which the PP *on the towel* was the destination for the apple (in other words, a VP-attached PP). However, in the presence of two possible referents, adults waited for disambiguating information about the referent prior to committing to a destination. In other words, the visual context influenced adults' parsing behavior, prompting them to wait for additional information rather than immediately committing to a specific interpretation.

In summary, adult sentence processing is notably incremental. These parsing decisions are made on the basis of lexical information and can be influenced by linguistic or visual context. In the next section, we cover the characteristics of early sentence processing in children.

5.2.2 Child sentence processing

Children seem to display many of the same characteristics as adults in their processing of sentences, but differ in important ways. In a seminal study investigating early sentence processing, Trueswell et al. (1999) extended the work of Tanenhaus et al. (1995) to young children. They prompted both 5-6 year olds and adults to act out temporarily ambiguous sentences such as (29):

(29) Put the frog on the napkin into the box.

while viewing one of two layouts. In the 1-referent condition (Figure 5.1, left), only one frog was present; in the 2-referent condition (Figure 5.1, right), two frogs were present. Just as in Tanenhaus et al. (1995), both children and adults glanced at the napkin in the bottom right corner when in the 1-referent condition. These glances suggest that both children and adults initially interpret the PP *on the napkin* as the destination for the frog, on the basis of the fact that *put* requires both an object and a destination. Unlike adults, however, children fail to revise from this initial interpretation: while adults correctly move the frog in the bottom left to the box in the top right upon hearing the rest of the sentence, the majority of children proceeded to move the frog to the napkin in the bottom right, acting out the initial interpretation. This behavior suggests that children are unable to revise from their initial parsing commitments.

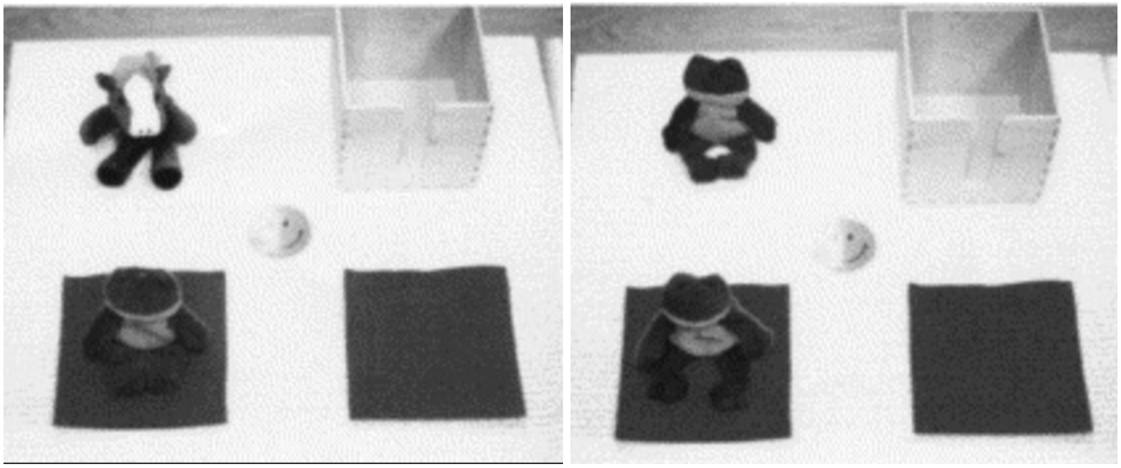


Figure 5.1: The 1-referent (left) and 2-referent (right) experimental layouts in Trueswell et al. (1999).

Children's behavior further differed from that of adults' in the 2-referent condition (Figure 5.1, right). In this condition, replicating the findings of Tanenhaus et al. (1995), adults made fewer visual fixations on the napkin in the bottom right quadrant, instead waiting for disambiguating cues for the referent. In other words, because there were two frogs in the scene, adults appeared to wait for additional information to make predictions about the destination for the frog until they knew which frog they were being asked to move. Children, on the other hand, showed a completely different pattern: just as in the 1-referent condition, children glanced at the napkin in the bottom right quadrant, suggesting that they were continuing to make early predictions regardless of the ambiguous referent. The majority of children then chose one of the two frogs (at chance) and moved them to the napkin in the bottom right corner. Crucially, children's performance with unambiguous sentences like (30) in both conditions is adult-like:

(30) Put the frog that's on the napkin onto the book.

suggesting differences in children's and adults' performance on this task stemmed from differences in processing strategies, rather than lack of linguistic knowledge of, or inability to parse, relative clause constructions.

Trueswell et al. (1999) provides some of the first evidence that children's online sentence processing is incremental, just as in adults. However, their findings also highlight the fact that children's sentence processing differs from that of adults in two key ways. First, children make early commitments during parsing and are unable to revise from these commitments. Second, although both children and adults use their knowledge of a verb's argument structure to guide their predictions and incremental interpretations (MacDonald, 1997; Snedeker & Trueswell, 2004; Tanenhaus et al., 1995; Trueswell et al., 1999), children do not appear to use context to inform their parsing strategy. In addition, enhanced pragmatic cues and emphasis on referential cues (such as the experimenter highlighting the fact that there are two frogs, one on a napkin and one on another surface) do not aid in the garden-path recovery of 5-year-olds (Weighall, 2008), although slower speech rates may allow children to recover at slightly higher rates (Qi et al., 2020).

Further work provides additional support for 5 year-old children relying heavily on verb bias to parse sentences. Snedeker and Trueswell (2004) instructed both adults and children to complete an action with sentences containing instrument PPs, in frames that contained verbs with either a 'modifier bias' (*choose*), 'equal bias' (*feel*), or an 'instrument bias' (*tickle*). For example, a participant could be asked to *Choose the dog with the pen* (where an initial bias would be to choose the dog that is holding a pen, rather than to use the pen to choose the dog), to *Tickle the dog* *with the pen* (where the bias would be to use the pen to tickle the dog), or to *Feel the dog with the pen* (where it is equally likely to interpret *the pen* as either a modifier or an instrument, and thus unbiased). In some cases, the visual scene contained two of the target animal, one with the item, and one without. Having two animals prompted adults to follow a modification-bias action even in the instrument-bias condition. For example, if given the choice between a dog holding a pen and a dog without a pen, an adult instructed to *Tickle the dog with the pen* would tickle the dog holding a pen with their finger, rather than using a pen as an instrument to tickle one of the dogs. However, changes to the visual context did not influence children's parsing decisions, which fell more in line with the biases of the verbs used in the instructions. This suggests that children rely heavily on verb biases to inform their parses. Moreover, this study provides further evidence that children cannot revise their initial interpretations, even in the face of conflicting referential information.

This seeming inability to revise has also been implicated in a host of other studies probing the developing sentence processor. For example, Omaki et al. (2014) tested 6 year-old children on ambiguous bi-clausal wh-questions like (31) and (32):

(31) Where did Lizzie tell someone that she was gonna catch butterflies?

(32) *Doko-de Yukiko-chan-wa choucho-o tsukamaeru-to itteta-no?*
where-at Yukiko-DIM-TOP-PRO butterfly-ACC catch-COMP telling-Q
"Where was Yukiko telling someone that she will catch butterflies?"

Children appear to initially interpret the linearly closest clause to be the one that the wh-filler is associated with (that is, in Japanese, children interpret the question to be about catching butterflies, while in English, children interpret it to be about the act of telling someone about catching butterflies). Adults initially make the same interpretations. Unlike children, however, adults are capable of revising these interpretations in the face of conflicting evidence, as in sentences that syntactically block one of the interpretations (33):

(33) *Doko-de Yukiko-chan-wa [*PRO *kouen-de choucho-o tsukameru-to] itteta-no?*
where-at Yukiko-DIM-TOP [she park-at butterfly-ACC catch-COMP] telling-Q
"Where was Yukiko telling someone that she would catch a butterfly at the park?"

Here, the sentence can only be answered with the location of telling someone about catching butterflies because the location of the embedded clause event is already specified (the park), blocking that interpretation. When presented with this type of sentence, however, children do not appear to be able to revise from the initial association (the location of catching the butterflies), suggesting once again that the child parser has difficulty with revision.

Evidence for the use of lexical information in predictive parsing is even present in infants. Lidz, White, and Baier (2017) tested 16 month-old infants and 19 month-old toddlers on sentences with nonce nouns, such as (34a) and (34b) while viewing a scene in which an actor wipes a toy camera with a cloth:

(34) a. She's wiping the *tiv*.
b. She's wiping with the *tiv*.

If children can identify that *tiv* is the direct object in (34a) and a prepositional object in (34b), and use that information to infer the likely thematic relation of the novel noun – patient in (34a), instrument in (34b) – they should look towards the corresponding instrument (cloth) or patient (toy camera) in the test trials. Interestingly, only 16 month-olds reflected successful word-learning, looking to the patient after hearing (34a) and looking to the instrument when hearing (34b), while 19 month-olds looked significantly more to the patient in *both* conditions.

One possibility is that the seeming inability of 19 month-olds to learn the novel noun in sentences like (34b) is an artifact of predictive and incremental parsing: at 19 months, the children are predicting a "Ving NP" structure (rather than the correct "Ving *with* NP" structure), using distributional information about the typical argument structure of *wipe* to make this prediction. These predictions, however, interfere with children's performance because of possible difficulties revising their initial parse, making them appear less competent than 16 month-olds. To test this hypothesis, the authors presented the 19 month-olds with sentences containing an additional referentially ambiguous expression, as in (35a) and (35b):

(35) a. She's wiping that thing with the *tiv*.

b. She's wiping the *tiv* with that thing.

Here, the 19 month-olds succeeded, looking more at the target instrument after hearing (35a). This supports the hypothesis that 19 month-olds are using their knowledge of the verb argument structure of *wipe* to predictively parse: in this case, the prediction that an NP will follow *wiping* is satisfied, allowing the 19 month-olds to correctly learn that *the tiv* is the instrument of wiping, while *that thing* is the patient.

In summary, the child parser differs from that of adults in a key way – namely, the seeming inability to revise from an initial interpretation. However, the child parser is similar to that of adults in that it is also incremental and predictive, and relies on knowledge of verb argument structure to guide predictive parsing. Two hypotheses have been proposed in the literature in an attempt to explain the parsing differences between children and adults. In the first hypothesis, a developing working memory system impedes the parser's ability to return to already processed information because it may have been "discarded" due to limited capacity, thereby precluding revision (Montgomery et al., 2008; Trueswell et al., 1999; Weighall & Altmann, 2011; Zhou et al., 2021). In the second hypothesis, a still-developing cognitive control system impedes the parser's ability to inhibit an initial parse, thus preventing revision even if alternatives have been maintained in memory (Hsu et al., 2021; Novick et al., 2005; Ovans, 2022; Woodard et al., 2016). To better understand the interaction of these two systems with early sentence processing, in the next section, we provide the necessary background on what is known about these two systems, with respect to language processing, and over the course of development.

5.3 Extralinguistic cognition and sentence processing

Adults' parsing difficulties with "garden-path" sentences suggest that the mature parser is subject to constraints from extralinguistic systems like working memory or executive control. These extralinguistic cognitive systems place limitations on the parser's ability to maintain or entertain all possible alternatives at any given choice point, leading to errors of comprehension. Consequently, understanding precisely how these systems are recruited during parsing is crucial for our understanding of how the parser develops.

5.3.1 Extralinguistic cognition and sentence processing in adults

5.3.1.1 Working memory and executive control in adults

The role of working memory in sentence processing is well-studied, with some of the most prominent models of sentence processing developed with the goal of minimizing memory load (e.g., Frazier, 1979). It is important to note that the working memory system we refer to here is broader than the visual working memory that was the focus of Part 1. While visual working memory constrains visual perception, the components of the working memory system that influence sentence processing pertain more to the ability to maintain and manipulate information used over the course of executing a larger cognitive task (such as maintaining alternative analyses for a sentence as it unfolds).

In humans, working memory is a component of the broader memory system specifically tasked with providing temporary storage of the information necessary for other, more cognitively taxing tasks. It is thought to serve as a bridge between perception, long-term memory, and action, facilitating completion of larger tasks by maintaining information about intermediary steps (Baddeley, 1992, 2003; Baddeley & Hitch, 1974; Just & Carpenter, 1992, among others). The exact structure, components, and role of working memory are a topic of hot debate within the cognitive psychology literature. Nonetheless, the field has reached a consensus that working memory is characterized by its limited capacity, which requires specialized attentional mechanisms and peripherally-based storage mechanisms to maintain relevant information in pursuit of completing a particular task. For the purposes of this dissertation, we abstract away from the debate of whether the working memory system is comprised of a single unitary system (e.g., Engle et al., 1992), or a multi-component system (e.g., Daneman & Carpenter, 1980; Daneman & Tardif, 2016). The question that will pertain most to the current study is not the precise architecture of the system, but rather the broader consensus that the capacity of this system is limited.

One of the most well-known models of the working memory system was proposed by Baddeley and Hitch (1974). They proposed that working memory consists of a three-component system, comprised of a phonological loop, a visuospatial sketchpad, and a central executive system which can modulate attention to/from these two storage systems. Within sentence processing, the phonological loop and the central executive are most relevant. The phonological loop consists of a phonological store, which maintains phonological information for a brief period of time (several seconds) before it fades. This phonological information can be 'refreshed' through an articulatory rehearsal process, in which the phonological information is repeated subvocally to maintain it in the system. This system is limited by the amount of time it takes to rehearse the phonological elements present: as the number of items being maintained increases, the first items will have faded by the time the articulatory rehearsal of the full list has been completed. The phonological loop is activated whenever a list of elements is being maintained by the system, and is therefore the target for most metrics of working memory capacity, including immediate serial recall tasks which require participants to maintain a set of digits, letters, or words. This phonological loop may also be recruited during sentence processing in the case the parser encounters a parsing error, allowing for rehearsal of the sentence and enabling revision of the parse if necessary.

Another relevant component Baddeley and Hitch (1974)'s working memory model is the central executive, which is responsible for the allocation of attentional resources between the items being maintained. This system was initially proposed as a resource of processing capacity. At least one account of the interaction of working memory and sentence processing (Just & Carpenter, 1992) expands on this capacity model, where the amount of attentional resources governs the capacity of the working memory system. We will discuss this model in more detail in the next section. Other accounts have focused more on the mechanism by which this attentional control is allocated. For example, Baddeley and colleagues adopted the Norman and Shallice (1986) model of attentional control for this component of working memory. This model of attentional control consists of two processes: one is responsible for the control of behavior through habits and ingrained action schemas, and the second is responsible for overriding this type of routine control when necessary. Thus, this model posits that two tasks can interfere with one another either by a) requiring the same processing structures, or b) competing for the same attentional resources. The Norman and Shallice (1986) model of attentional control is one of the first explicit models of cognitive control, another crucial component of extralinguistic cognition which could be recruited during sentence processing.

Cognitive, or executive, control is involved in monitoring and resolving cognitive conflicts when they arise (Botvinick et al., 2001; Milham et al., 2001; Miller, 2000; Miller & Cohen, 2001; Ye & Zhou, 2009). Conflicts are caused by the incompatibility of multiple representations (such as visual or linguistic representations) or the opposition of action tendencies (such as conflicting habits and required actions). Within sentence processing, a representational conflict could arise if two incompatible linguistic analyses compete, as in the case of a temporarily ambiguous sentence where an initial prediction may be incompatible with later arriving information. Within the cognitive control literature, a classic example of representational conflict is the Stroop task, in which participants are presented with color words printed in different colors, and asked to respond with the print color (Stroop, 1935). These representations are either congruous, (e.g., the word "green" with a green font color) or incongruous (e.g., the word "green" with a red font color). Incongruent trials require the participant to suppress the immediate representation of the orthographic information (reading "green" is automatic) in favor of the task-relevant goal of identifying and naming the font color. This incongruency results in slower response times and increased errors in participants.

The suppression of task-irrelevant representations is thought to be achieved through a biasing mechanism, which promotes goal-relevant information over goal-irrelevant information (Botvinick et al., 2001; Braver, 2012; Cohen & Huston, 1994; Milham et al., 2001; Ness et al., 2022). Although the exact mechanism by which this biasing occurs is the subject of debate, Botvinick et al. (2001) proposed a model in which representations are formed for both units of information producing the conflict. Both representations are activated, resulting in conflict. As two units are concurrently activated, they increase the activation of the conflict monitoring node of the cognitive control system, which then executes top-down control by increasing the weight of the unit corresponding to the goal of the task.

The ability to recruit cognitive control in order to suppress irrelevant information is crucial for sentence processing, as language users regularly encounter conflicting linguistic representations. For example, an early parsing commitment may be incompatible with late-arriving bottom-up information, resulting in conflict between the early parse commitment and the parse supported by the bottom-up information. We now turn to how the extralinguistic systems of working memory and cognitive control are recruited during parsing.

5.3.1.2 Extralinguistic cognition and sentence processing

The working memory system is thought to be recruited during sentence processing to facilitate the storage of intermediate and final analyses as a comprehender incrementally integrates novel words and information. Just and Carpenter (1992) proposed a capacity-based model of how working memory constrains language comprehension. In this model, both storage and processing are guided by activation, and the capacity of the working memory system is the maximum amount of activation that is available to support either storage or processing. Each element of processing – whether it is a particular word, a syntactic structure, a thematic structure, an aspect of real-world context, or other – takes up some of this capacity so long as its activation is above some minimum threshold. Once the system's capacity is full, however, the activation for the oldest elements is reduced, resulting in a "loss" of those elements. Thus, if a particular sentence requires the comprehender to maintain some aspect of information for an extended period (e.g., temporary ambiguity), they may exceed the capacity of their system and lose this information if understanding the sentence also requires additional manipulation.

Evidence for this capacity-based model of working memory comes from correlations of working memory measures with reading times on sentences with reduced relative clauses. Using the Reading Span task (Daneman & Carpenter, 1980), which requires participants to listen/read to a series of sentences and remember the last word used in that sentence, Just and Carpenter (1992) found that higher reading spans (more words remembered) correlated with faster ambiguity resolution in sentences with reduced relative clauses. In other words, individuals able to retain more information in their working memory (high-span) appear to have an advantage in parsing sentences with reduced relative clauses that are temporarily ambiguous with respect to their attachment.

Just and Carpenter (1992) also found that high-span readers were better equipped to utilize the animacy of the grammatical subject to resolve ambiguity. Because inanimate nouns are unlikely to be the agents of a verb, animacy can be a useful pragmatic cue indicative of the presence of a reduced object-relative clause. For example, in the sentence *The defendant examined by the lawyer shocked the jury*, the animate subject may lead readers “down the garden path” to believe *The defendent* is the agent of *examining*. On the other hand, in the sentence *The evidence examined by the lawyer shocked the jury*, readers able to exploit animacy as a pragmatic cue could realize that *The evidence* is unlikely to be the agent and thus make fewer parsing errors. While low-span readers showed longer reading times for reduced relative clause sentences with both animate and inanimate subjects, high-span readers had reduced reading times for sentences with inanimate subjects. This suggests that individuals with higher working memory capacity may be more sensitive to the pragmatic cue of animacy. Just and Carpenter (1992) argue that these findings point to a finite resource of working memory: individuals with larger capacities (high-span group) are able to incorporate pragmatic cues, while those with smaller capacities (low-span group) cannot due to the limit on working memory. Additional studies have confirmed that working memory capacity, as measured by linguistic working memory tasks such as the Reading Span task, has strong correlations with the efficiency of sentence processing during reading comprehension (Caplan & Waters, 1999; Lewis, 1996; O'Rourke, 2013).

Others have argued that it is the interaction of working memory with executive control that is responsible for these effects. Using another task typically associated with working memory, Novick et al. (2014) trained participants on an *N*-back test, in which participants are presented with sequences of single letters and asked to judge if the current letter is the same as one displayed *N* letters prior. For example, a participant in a 4-back task would be shown a sequence of letters like X P U X X̲ P P, and would then be asked to identify whether the letter currently displayed (the first "X") matched the letter four places prior (the underlined "X"). This task also includes so called *lures*, which occur in the *N*±1 places (the "X" immediately preceding the underlined "X"). Novick et al. (2014) found that training on this type of *N*-back task can improve participants' recovery from initial misinterpretations of garden-path sentences relative to their performance before training. Although the improvement of processing after training the *N*-back task could be viewed as evidence for the role of working memory in parsing, Novick et al. (2014) found no correlation with improved performance on *N*-back trials on sentence comprehension. Instead, only training on the detection of *N*±1 lures improved participants' ability to recover from garden-path sentences. Because the presence of lures raises a conflict between the goal at hand (maintaining the item *N* trials prior), Novick et al. (2014) argue that improved performance is not evidence of working memory's role in sentence processing, but rather evidence of cognitive control engagement.

Recruitment of cognitive control may be critical for sentence processing, given the multitudes of linguistic representations that must compete with each other over the course of parsing (Novick et al., 2005). For example, in sentences with global ambiguity, two plausible interpretations can be accessed (e.g., *Touch the dog with the pen*, where the PP *with the pen* could either be VP-attached or NP-attached). Similarly, temporarily ambiguous sentences such as *Put the frog on the napkin in the box* require the listener to suppress the initially-plausible interpretation that the napkin is the destination for the frog (VP-attachment) once additional information arrives and it becomes evident that the PP *on the napkin* modifies *the frog* instead.

Evidence that cognitive control is recruited during sentence processing comes from a variety of sources. Findings from neuroimaging studies have highlighted that the same brain region activated during nonlinguistic conflict resolution (the left inferior frontal gyrus, LIFG) is also activated during linguistic conflict (Hsu & Novick, 2016; January et al., 2009; Ye & Zhou, 2009). Similarly, damage to the LIFG appears to impair real-time language processing decisions. Although patients with lesions to the LIFG appear to maintain general language skills (performing well on aphasia test batteries), they fail to recover from misinterpretations of syntactically ambiguous input (Novick et al., 2009). Lastly, engagement of cognitive control in nonlinguistic tasks appears to facilitate listeners' incremental processing of temporarily ambiguous spoken instructions, decreasing the number of initial misinterpretations. Hsu and Novick (2016) interleaved trials of the Stroop task (Stroop, 1935) with language comprehension tasks involving syntactic ambiguity (i.e., acting out the sentence *Put the frog on the napkin in the box*). They found that participants' rate of incorrect actions was significantly reduced following incongrous Stroop trials compared to congrous Stroop trials, suggesting that the activation of cognitive control can improve real-time sentence comprehension. This improvement was later confirmed not to be the result of upregulated attention, and indeed the result of upregulated cognitive control (Hsu et al., 2021).

Given the wealth of evidence that working memory and cognitive control influence sentence processing in adults, it is plausible that these systems under development could be the source of the parsing difficulties in children discussed in §5.2.2. In the next section, we discuss the development of these two extralinguistic systems, as well as their possible influence on early sentence processing.

5.3.2 Extralinguistic cognition and sentence processing in children

5.3.2.1 Working memory and cognitive control in development

It may come as no surprise that children's cognitive capacities are limited early in childhood and develop considerably as they mature. Working memory is no exception: investigations of working memory across development have shown that performance on complex span tasks improves during childhood (Case et al., 1982; Towse et al., 1998). Although children's spans are typically smaller than those of adults, they appear to rely on a similar architecture as adults. Bayliss et al. (2005) found that children also make use of a phonological loop, and are able to repeat up to 1-3 syllable items, although performance decreased at and above 4 syllables. Working memory performance in children appears to be modulated largely by processing speed, with rate of recall (the speed with which a set of items held in memory, such as letters, numbers, or a sequence of taps, can be recalled) and speed of search (the rate with which a child "searches" for targets in a visuospatial memory task, measured by the number of taps to a screen) most correlated with working memory span (Chuah & Maybery, 1999). However, storage constraints appear to be domain specific: maintaining the same types of elements in memory cause more storage difficulty than storing different items (Bayliss et al., 2005; Bayliss et al., 2003).

Cognitive control also takes a notoriously long time to develop in children. Typically developing children between the ages of 3 and 6 demonstrate high levels of processing difficulty and interference on cognitive control tasks such as the Day/Night task, in which children are asked to refer to an item by its antonym (e.g., calling the night "day", Anderson, 2002; Diamond et al., 2002; Gerstadt et al., 1994) and age-appropriate Stroop tasks (Prevor & Diamond, 2005). Similarly, children's performance on a dimensional change card sort task points to an immature cognitive control system. In this task, children are presented with cards depicting items with features varying on two dimensions (e.g., color and shape) and are then asked to sort cards by one dimension (i.e., on color). In the second phase of this task, the sorting dimension is later switched (i.e., to shape). Children between 3 and 4 years old 'perseverate' and continue to sort by first dimension, even when reminded of the new rules, while most 5-6 year olds can switch successfully (Doebel & Zelazo, 2015). Children's performance on these tasks indicate an immature cognitive control system which impairs children's ability to suppress prepotent responses.

5.3.2.2 Interactions of extralinguistic cognition with sentence processing in children

Developing working memory and cognitive control have both been implicated in children's inability to revise their initial predictions, as well as their over-reliance on lexical bias to guide their parses. Recall from §5.2.2 that children's performance with temporarily ambiguous sentences like *Put the frog on the napkin in the box* is non-adultlike in that children carry out an action on the basis of their initial interpretation, moving the frog to another napkin rather than into a box (Trueswell et al., 1999). This is thought to occur as a result of children's use of the knowledge of the selectional properties of *put*, which requires an object to be moved and a destination for that object, to inform early parsing commitments.

One possible explanation of children's parsing difficulties in Trueswell et al. (1999)'s study is that children's working memory is limited. As a result, children pursue the initially preferred parse while less likely alternatives (such as a *put* sentence with two PPs) are discarded due to limited memory capacity. Once this alternative has been "lost", the initial interpretation cannot be revised. In contrast, adults have larger working memory capacities, allowing more alternatives to be retained and leaving more interpretations to revise to. Working memory performance has been correlated with recovery from garden path effects in children (Weighall & Altmann, 2011; Zhou et al., 2021). More broadly, working memory span has been correlated with comprehension time in complex sentences for children between the ages of 6-12 years old (Montgomery et al., 2008), suggesting that developing working memory can influence early sentence processing and comprehension.

Another possible explanation for children's performance with temporarily ambiguous sentences in Trueswell et al. (1999) is that an immature cognitive control system impedes the ability to successfully handle situations of "representational conflict" – that is to say, late-arriving linguistic evidence incompatible with an earlier parsing commitment. Cognitive control would thus fail to be sufficiently engaged when the system needs to disregard the preferred (or most probable) analysis in favor of an initially less-preferred option to arrive at the correct interpretation. In other words, to successfully parse the sentence *Put the frog on the napkin in the box*, children would need to suppress the initially-plausible interpretation that *on the napkin* is the destination for the frog in order to pursue the correct parse, where it modifies *the frog*, and *in the box* is the true destination. Indeed, Woodard and colleagues (2016) found correlations between children's garden path recovery and 3 different measures of their cognitive control, but did not find a correlation with garden path recovery and performance on a linguistic working memory task. Interestingly, unlike in adults, activating cognitive control in earlier tasks may not improve recovery on garden path sentences: Ovans (2022) found that children's performance on difficult sentences immediately following incongruous Stroop trials decreased (c.f., adults whose performance is increased following incongruous trials, Hsu et al., 2021).

Both immature working memory and cognitive control systems have been implicated in children's early parsing difficulties. However, findings linking the two systems to parsing performance are almost exclusively correlational, thus making it difficult to determine precisely how each developing system may be influencing early sentence processing. In order to better examine the interaction of these two systems with sentence processing throughout development, in Chapter 6 we develop a computational model that makes explicit the roles of both working memory and inhibitory control in the developing parser. In formally paramaterizing the roles of each of these two systems in our computational model, we are able to individually probe the role of each system, allowing us to generate testable predictions for the interaction of each immature system with sentence processing, which can be examined in future experimental work. We test our approach on the sentences from Trueswell et al. (1999).

5.4 Summary

Human sentence processing is incremental, predictive, and lexically-dependent in both adults and children. However, children differ from adults in their ability to update their commitments and revise their initial parsing commitments. In particular, children appear to rely heavily upon lexical bias, using their knowledge of verb argument structure to guide their early commitments to parses. They also appear to be unable to revise if they encounter late-arriving information incompatible with that early commitment. The source of children's inability to revise from these commitments is still unknown, although it has been hypothesized that developing extralinguistic cognitive systems may be responsible. Two such extralinguistic cognitive systems have been correlated with sentence processing in both children and adults: working memory and cognitive control. Because these systems are still developing in children, it has been hypothesized that one or both of these systems is responsible for children's inability to revise from their initial commitments.

However, the respective roles of working memory and cognitive control are difficult to tease apart experimentally, and evidence for the influence of these systems on early sentence processing has been exclusively correlational. Another approach is needed to tease apart the influence of each of these systems. Namely, a formalized model of each system's role in sentence processing over development is necessary to pinpoint the mechanisms by which either working memory or cognitive control interact with sentence processing in children. In the next section, we develop a formal computational model of the interaction of cognitive control and working memory on sentence processing in both children and adults. To this end, we focus on capturing the empirical findings of the seminal Trueswell et al. (1999) paper, with the goal of eventually extending this model to other empirical findings in future work.

CHAPTER 6

Computational model of sentence processing in development

In this chapter, we implement a generalized left-corner parser with parameters corresponding to developing working memory and cognitive control and report on performance with the well-known example from Trueswell et al. (1999). As discussed in Chapter 5, the difference between children and adults' early parsing performance has been attributed to either limited working memory, immature cognitive control, or both. A computational model with different parameters corresponding to either working memory or cognitive control limitations allows for adjudication between these systems as the cause for children's seeming inability to revise from early parsing commitments. To successfully model parsing across development, the parser should be able to accommodate adult-like parsing at one setting of a parameter (for example, greater capacity in working memory), and also accommodate child-like parsing at a different setting of that parameter (for example, smaller capacity in working memory). We implement such an incremental and predictive computational parser that parameterizes the roles of working memory and cognitive control. These parameters can be individually and incrementally adjusted explore the predictions of hypothesized changes in these systems.

6.1 Previous models of extralinguistic cognition and ambiguity processing

Although computational models of the garden path effect (and other parsing difficulties) have been implemented for adults (e.g., Brants & Crocker, 2000; Crocker & Brants, 2000; Jurafsky, 1996), these models do not explicitly accommodate the developing parser. In this section, we examine existing models that have taken into account working memory and cognitive control. We expand on these models in our integrated model of parsing in development in the remainder of this chapter.

Jurafsky (1996) proposed a parsing model which begins with probabilistic lexical access, generating linguistic structures in a bottom-up manner. These structures are then disambiguated by "pruning", or abandoning, unlikely parses. This pruning occurs through a beam search algorithm. Beam search algorithms, originally developed for optimized natural language processing (NLP), explore a hypothesis space by expanding the most promising node in a limited set (Lowerre & Reddy, 1976). Beam search uses a priority queue of candidate analyses: candidates are ranked by scores based on their relative probabilities. In a k-best beam search, all alternatives are weighed against the highest scoring option. If the scores of these alternatives fall within the prespecified threshold k (where k is a ratio of parse scores), they are pursued, while all other options "fall off" the beam and cannot be pursued. Jurafsky (1996) proposes this threshold k as a working memory limitation: if a given analysis is not "activated" sufficiently and does not have a high enough score, it falls below the retention threshold and is "lost" from the working memory store (Just & Carpenter, 1992). Jurafsky (1996) argues this "loss" could explain garden path effects and other processing errors.

With k-best beam restriction, Jurafsky (1996)'s model accommodates several types of garden path effects. For example, when set to a k of 0.2, the model captures empirical data on parsing difficulties, such as slower reading times in sentences like (36a), compared to sentences like (36b):

(36) a. The complex houses married and single students and their families. (Hearst, 1991)

b. This complex houses married and single students and their families.

This example is thought to cause difficulty for adults because most readers initially interpret 'The complex houses' as a noun phrase rather than the intended noun phrase ('The complex') and verb ('houses'); the sentence in (36b) bars this interpretation. Jurafsky (1996)'s parser models this difficulty by 'pruning' the analysis in which *The complex* constitutes an NP and *houses* heads a VP, because it is structurally dispreferred and has a lower probability compared to the analysis in which *The complex houses* constitutes the NP. This latter analysis is maintained due to its higher probability overall (an analysis with an NP containing a determiner, adjective, and noun is more probable than an analysis where *houses* is a verb). As a result, when the parser continues parsing the string, it is led to a parsing error, analogous to a garden path effect, when the rest of the sentence cannot be accommodated in this structure. Other computational models have also used similar variable-width beams to model memory limitations, only maintaining the analyses that meet some prespecified threshold *k* (Brants & Crocker, 2000).

While Jurafsky (1996)'s model formalizes the influence of limited working memory on sentence processing, other models of sentence processing incorporate elements of cognitive control. For example, although not a formal computational model, MacDonald et al. (1994)'s model of constraint-based human sentence processing relies on excitatory and inhibitory connections between representations. For example, linguistic representations of specific syntactic features, such as voice (e.g., active or passive) or argument structure (e.g., transitive or intransitive) exist in an inhibitory relationship with each other. Thus, in the sentence *The defendant examined by the lawyer shocked the jury*, the verb *examined* activates both active and passive voice nodes. In order to make parsing the sentence possible, these nodes exert inhibitory force on each other. In this case, the active voice node exerts greater inhibitory force due to the fact that *The defendant* is an animate subject and thus more likely to be an agent of *examining*. In this way, the parser is biased towards the active voice node. However, by inhibiting the passive voice node, the reader is lead "down the garden path" – *the defendant* is actually the patient of examination by *the lawyer*. With these inhibitory relationships between nodes, MacDonald et al. (1994)'s model nods towards a cognitive control system. However, this model does not explicitly recruit extralinguistic inhibitory control in the case of conflicting evidence. Instead, the model proposes that upon encountering this error, the linguistic context could lead to more activation of the passive node. Similarly, other models of sentence processing have cited such inhibitory and excitatory connections (e.g., Vosse & Kempen, 2000). Although these models suggest that conflict could force the parser to redistribute the activation of different nodes, they do not formally propose a mechanism for the reallocation of activation between nodes. Thus, a formal model of how cognitive control influences sentence processing is still needed.

In the next two sections, we describe our approach to modeling both cognitive control and working memory in the developing parser. Our parser individually parameterizes each of these systems. These parameters can be adjusted independently, which allows us to test two hypotheses about which system is primarily responsible for the observed differences between child and adult sentence processing. Adjusting each of these parameters reflects maturation of that cognitive system, further allowing us to model parsing at multiple stages of development. We begin by describing the general parser architecture in §6.2. In §6.3, we describe the two parameters that correspond to cognitive control and working memory.

6.2 General parser architecture

We developed an incremental and predictive parser which implements a generalized left-corner strategy, using both top-down and bottom-up parsing strategies. In a generalized left-corner parser, building structure must be "triggered" by the bottom-up identification of the 'left corner' (for example, in the rule $A \rightarrow B\ C$, B might be the left corner). Once B is found, C will become a 'sought' constituent. We denote found constituents as X and predicted or sought constituents as $\overline{X}$.

The parser takes a probabilistic context-free grammar (PCFG) and a string, and returns zero or more parses for that string. Within the parser, a candidate analysis, $C_n = (D, S, ws_n)$, consists of a derivation D, a category stack S of found and sought constituents, and the string ws_n which remains to be parsed. The derivation D consists of the relevant sequence of rules from G that have been used in constructing the parse so far – essentially, the structure of the sentence that the parser has constructed. The stack S consists of non-terminal symbols that constitute 'loose ends', which are nodes that have either been found bottom-up and need to be integrated into a connected structure (X), or that have been predicted top-down and still need to be found in the string ($\overline{X}$). When written horizontally, the top of the stack is on the left.

Our parser pursues parsing decisions based on the relative probabilities of structures in child-directed speech. These probabilities are formalized as rule probabilities within a PCFG. A PCFG $G = (N, T, R, S^{\dagger})$, consists of a set of nonterminal symbols, *N*, a set of terminal symbols *T*, a start symbol $S^{\dagger}$, and a set of rules, *R*, of the form $(A \rightarrow \alpha, p)$, where *p* is the probability of the right-hand side α given the left-hand side *A* of the rule. This parser assumes a PCFG with only two types of rules: one set of rules (preterminal rules) introduces a single terminal symbol, and another set of rules introduces one or more non-terminal symbols. No empty strings can be the production of any rule.

Our model relies on a partially-lexicalized PCFG generated based on a large repository of child-directed sentences from the CHILDES Treebank (Pearl & Sprouse, 2013). A partially-lexicalized grammar allows some lexical information to be accessible in non-terminal nodes. We chose to use a partially-lexicalized grammar because we know that children (and adults) rely heavily on verb bias and their knowledge of argument structure to inform their parses. This allows for different probabilities for expansions of different verb phrases to relate to the lexical identity of the verb – for example, the probability of *put* having a *PP* argument is much higher than the probability of *wipe* having a *PP* argument. Encoding this into the grammar through partial lexicalization allows us to reflect what adults and children know about the argument structures of these verbs. The probabilities in the grammar were determined based on the distribution of parse trees of child-directed speech in the CHILDES Treebank (Pearl & Sprouse, 2013) grammar. The grammar used in Simulations 1-2 can be found in Table 6.3 in §6.4, where it is also described in more detail.

6.2.1 Transitions

There are four possible transitions for the parser, which are summarized in Table 6.1. For preterminal rules, the parser can either SHIFT or MATCH. If there is a rule in the grammar such that $A \rightarrow x_i$, the parser can SHIFT by adding the preterminal category A to the stack when it sees x_i as the next input symbol (e.g., if there is a rule in the grammar such that $N \rightarrow$ *'frog'*, this would entail adding *N* to the stack). Similarly, if there is a rule in the grammar such that $A \rightarrow x_i$, the parser can MATCH if the top category on the stack is a top-down predicted category that can rewrite as the next input symbol, i.e., $\overline{A}$. In other words, if a previous operation has predicted the preterminal symbol top-down, the parser can remove that symbol from the stack if the next word matches that category (e.g., if $\overline{N}$ is the top symbol on the stack, *'frog'* is the next input word, and there is a rule in grammar *G* such that $N \rightarrow$ *'frog'*, the parser can "match" these symbols and remove $\overline{N}$ from the stack).

Transition	*Candidate*	*Resulting candidate*	*Rule*
SHIFT	$(D_i, \Phi, x_i x_{i+1}...x_n)$	$(D_{i+1}, A\Phi, x_{i+1}...x_n)$	$A \rightarrow x_i$
MATCH	$(D_i, \overline{A}\Phi, x_i x_{i+1}...x_n)$	$(D_{i+1}, \Phi, x_{i+1}...x_n)$	$A \rightarrow x_i$
PREDICT	$(D_i, B_1\Phi, x_i...x_n)$	$(D_{i+1}, \overline{B_2}...\overline{B_m}A\Phi, x_i...x_n)$	$A \rightarrow B_1...B_m$
CONNECT	$(D_i, B_1\overline{A}\Phi, x_i...x_n)$	$(D_{i+1}, \overline{B_2}...\overline{B_m}\Phi, x_i...x_n)$	$A \rightarrow B_1...B_m$

Table 6.1: Parser transitions and the resulting changes to the stack *S* and input string *ws*. Φ denotes the remainder of the stack.

The remaining two actions are PREDICT and CONNECT (Figure 6.1). In a PREDICT, if there exists a rule in the grammar such that $A \rightarrow B_1...B_n$ and the parser has 'found' the corner (e.g., B_1 is on top of the stack) bottom-up (e.g., through SHIFTing), it can predict the parent and sister nodes (Figure 6.1A) by removing the corner B_1 and replacing it with $\overline{B_2}...\overline{B_n}A$. For example, if the top symbol on the stack is *P* and there exists a rule *PP* $\rightarrow$ *P NP* in the grammar, the parser can "predict" a *PP* and the remaining daughters by removing the *P* and adding $\overline{NP}$ *PP* to the stack. The parent is given a bottom-up symbol because if the predicted daughters are also found bottom-up, the entire resulting structure will have been found bottom-up. A CONNECT can only occur if there exists a rule in the grammar such that $A \rightarrow B_1...B_n$, the corner has been found bottom-up, and the parent node has been predicted top-down (Figure 6.1B). A CONNECT removes the corner and predicted parent from the stack ($B_1\overline{A}$) and replaces them with the $\overline{B_2}...\overline{B_n}$. For example, if the current configuration of the stack is *P*, $\overline{PP}$, the parser can "connect" the *P* to the structure with the rule *PP* $\rightarrow$ *P NP* by removing both of these symbols from the stack and adding top-down predicted $\overline{NP}$ to the stack.

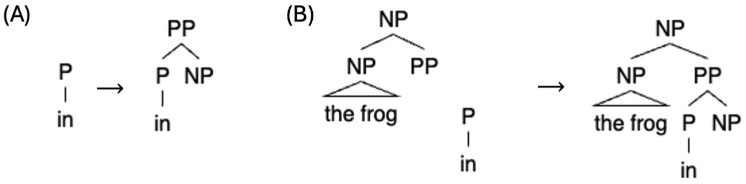


Figure 6.1: PREDICT (A, left) and CONNECT (B, right) in a left corner parser

The parser begins with the sought goal category ($\overline{ROOT}$) on the stack *S*. At each iteration of search, the parser identifies all of the actions that are possible in a given configuration, with each action resulting in a new configuration that is scored. We will now discuss candidate scoring.

6.2.2 Candidate scoring

For each configuration of the parser, the probability *p* of the candidate is the product of the top-down probabilities for each rule that has been used in the derivation. The product of these top-down rule probabilities is not conditioned on the left corner or goal category, which would yield the probability of taking a particular transition. Instead, *p* gives us the probability of the partial analysis that results from taking that transition. The probability *p* is used to calculate the score *s* for the given analysis. This overall score is the product of *p* and a lookahead probability *l* (discussed in the next section), taken to the (n+1)th root, where *n* is the number of connected nodes determined by the derivation constructed thus far (37):

$$(37) \quad s(D_1...Dn, S, w_1...w_n) = \sqrt[(n+1)]{((p(D_1) * ... * p(D_n)) * l(S, w_1))}$$

Taking this root allows for the normalization of configuration scores, preventing the unintentional penalization for deeper and larger trees that would occur if only the probability score was used (deeper trees have applied more rules, meaning more probabilities have been multiplied together and the score will necessarily be lower). Additionally, normalizing the score by the number of connected nodes penalizes disconnected structure – if the parser simply SHIFTs all of the input string and never CONNECTs or PREDICTs, the resulting score would be lower than a more connected structure by virtue of the smaller number of connected nodes n. Connected nodes are counted by adding the number of connections that are generated as a result of the transition. For SHIFT, 0 connected nodes are added because no connections have been made; for MATCH, 1 connected node is added because the found constituent has been connected to a sought constituent. For PREDICT, the number of connected nodes added is the length of the corner that has been found bottom-up (e.g., 1 if the corner consists of a single category). Lastly, for CONNECT, the number of connected nodes added is the length of the corner plus one, because a connection to the predicted structure is also formed.

6.2.2.1 Lookahead probability and PCFG announce points

We implement a one-word lookahead, which has been shown to significantly improve parser performance (Henderson, 2004). Lookahead implements the psychologically plausible possibility that there may be a brief lag between hearing a word and integrating it into the parse structure, during which time the system could have heard and recognized the next word in the speech stream. In essence, a one-word lookahead allows the parser to “peek” at the next word in the string and score candidate configurations according to whether or not they can generate that word.

We adapted two existing algorithms from Henriksen et al. (2019) and Nowak and Cotterell (2023) to calculate lookahead probability. If the symbol on top of the stack was top-down predicted, a one-word prefix probability was calculated using the algorithm from Nowak and Cotterell (2023). A one-word prefix probability is the probability that a string derived from a given nonterminal symbol begins with a given word. If the parser is in a configuration such that $\overline{NP}$ is the top symbol of the stack, and *the* is the next input string, the probability that an *NP* could yield *the* as the first symbol across the whole PCFG would be used as the lookahead probability. This aspect of the lookahead calculation is equivalent to that of Roark (2001).

If, on the other hand, the top symbol on the category stack is not a top-down symbol, lookahead was calculated by generating a derivative grammar as per Henriksen et al. (2019). Given a grammar with nonterminals A and B, the algorithm in Henriksen et al. (2019) constructs a new grammar with a nonterminal A_B, such that the strings derivable from A_B are exactly the strings that can be put after a string derivable from B to form a string derivable from A. In our parser, if the parser was in a configuration such that $P, \overline{PP}$ is on top of the stack, a derivative grammar would be generated with a nonterminal PP_P, in which the strings derivable from this nonterminal are exactly the strings that can follow P to form a string derivable from PP. The parser would then be able to use this nonterminal to calculate the probability of the next word (for example, *the*) given that a PP has been predicted and a P has been found bottom-up. In this way, the derivative grammar is used to generate the lookahead probabilities as in Nowak and Cotterell (2023) for stacks in which the first element is not a top-down prediction.

Lookahead is particularly important in the case of adjunction, where it may be beneficial to postpone the prediction of an adjunct until evidence for an adjunct has been found. This delay before connecting or predicting structure is in line with the experimental processing literature, which posits that adults treat adjuncts and arguments differently (Tutunjian & Boland, 2008) and wait to commit to a particular analysis until they encounter the first word of an adjunct (Sturt & Lombardo, 2005). In order to have this desired effect, we combine one-word lookahead with generalized left-corner parsing (Nederhof, 1993). To this end, announce points are set after the head of the NP in our PCFG. In a typical left-corner parser, the announce point is 1 (as illustrated in Table 6.1), meaning only a single constituent needs to be found to trigger the building of structure. In our PCFG, the announce point was set to 1 for all rules except for $NP \rightarrow D\ N$ and $NP' \rightarrow D\ N$, where the announce point was set to 2. In other words, both D and N must be found bottom-up prior to building structure. In contrast, if the announce point was set to 1, only the D would need to be found bottom-up prior to building structure. The addition of the second N constituent to the corner is especially important in combination with other NP rules in our grammar that have adjuncts, such as $NP \rightarrow NP\ PP$ and $NP \rightarrow NP\ SBAR$. By setting the announce point of $NP \rightarrow D\ N$ to 2 and implementing the one-word lookahead, the score of PREDICTing or CONNECTing to one of these rules with adjuncts will depend on whether or not the next word in the string signals an adjunct (e.g., if the next word in the input string is *that*, this will increase the probability of applying the rule *NP* → *NP SBAR* compared to other rules). Thus, setting the announce point to 2, in combination with one-word lookahead, allows the parser to postpone predicting an adjunct to an NP until it has seen the first word of the adjunct. If the announce point were 1, this prediction would necessarily come earlier, upon finding the determiner, because this would trigger the NP rule and prompt the parser to build structure ahead of seeing the adjunct. By combining this type of generalized left-corner parsing with one-word lookahead, we can implement the cognitively-plausible assumption that a listener waits until they hear the first word of an adjunct prior to hypothesizing the presence of an adjunct.

6.2.3 Beam search

The parser starts with candidate C_0 (in this case, the starting node $\overline{\text{ROOT}}$) on the beam. The parser then SHIFTs the first input symbol in the sentence. Candidates are assigned a score s, as per (37), based on the probability of the expansion and its lookahead probability l, normalized by the number of connected nodes in the structure so far. After scoring, the candidates are placed on a priority queue ranked by their score. The top-scoring configuration (or configurations, if scores are tied), $C_n = (D, S, ws_n)$, is popped from the priority queue. If ws_n = [ ] and S = [ ] (that is, all of the words in the sentence have been matched to a corresponding structure with the start symbol *ROOT* at its root), then the analysis is complete. Otherwise, all C' such that C_n derives C' are pushed onto the priority queue – all of the possible next steps are pushed onto a queue, ordered by their score, provided they meet some threshold k. This threshold is calculated by determining the top-scoring partial analysis C_{best} on the beam and in the candidate next steps and comparing each partial analysis in the next steps and the beam to that score. If the ratio of the scores between the candidate analysis and the highest scoring analysis is greater than k, the analysis is pushed to the beam. An analysis is discarded if $s_{Cn}/s_{Cbest} < k$. Then, the process repeats, popping the top-ranked candidate analysis (or analyses) again until no more candidates remain. Here, k must be within the range of 0 to 1, since it is a ratio of two candidate scores. At k = 1, the parser essentially becomes a serial parser: only one analysis can be pursued at any choice point. At *k* = 0, all grammatically-licensed analyses are found.

Following Jurafsky (1996), we implement a *k* threshold as a way of implementing bounds on working memory, which are assumed under any theory of human parsing (Frazier, 1979; Jurafsky, 1996; MacDonald et al., 1994, among others). This threshold naturally gives rise to a way of modeling memory development, as we will explain in the next section.

6.3 Modeling developing extralinguistic cognition

In this section, we outline our implementation of two parameters corresponding to the effects of working memory and cognitive control on parsing. These parameters can be independently adjusted to model the development of each system as the parser "matures" and approaches adult-like states of working memory and cognitive control as they relate to sentence processing.

6.3.1 Working memory

We model developing working memory through the adjustment of the parameter *k* in the *k*-best beam search pruning algorithm (Jurafsky, 1996). By adjusting the parameter *k*, we mimic a developing working memory system by increasing or decreasing the threshold against which candidate score ratios are compared: the only analyses pursued are those that have a high enough probability compared to the highest-ranked parse. Any parses with very low probability (as compared to the highest-ranking parse) fall out of the search beam and are consequently not explored. It follows that if *k* is set to high values, only candidates that score very closely to each other at each choice point will be pursued. If, on the other hand, *k* is set to low values, even candidates that differ significantly in their scores may be pursued (provided the ratio of the two scores is still above *k*).

Adjusting the *k* threshold could model the kindergarten path effects in Trueswell et al. (1999). Recall that in this study, children fail to understand the temporarily-ambiguous sentence (38a) but succeed in understanding the unambiguous sentence (38b):

(38) a. Put the frog on the napkin in the box Temporarily-ambiguous

b. Put the frog *that is* on the napkin in the box Unambiguous

Children appear to make an early parsing commitment to an analysis in which *the napkin* is the destination for the frog, which is consistent with a structure like Figure 6.2, left. However, once *in the box* is encountered, children seem to be unable to revise their initial commitment to pursue the structure corresponding to the intended interpretation of the sentence (Figure 6.2, right).

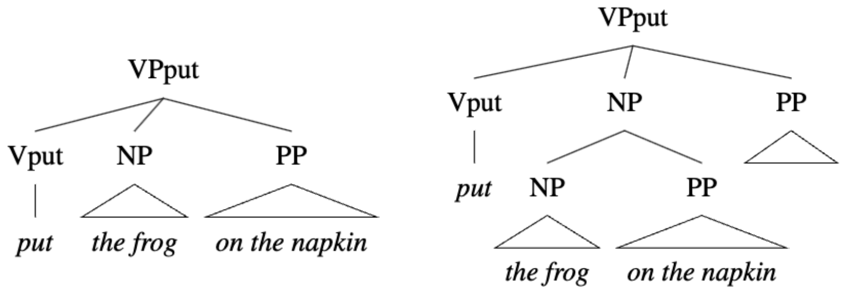
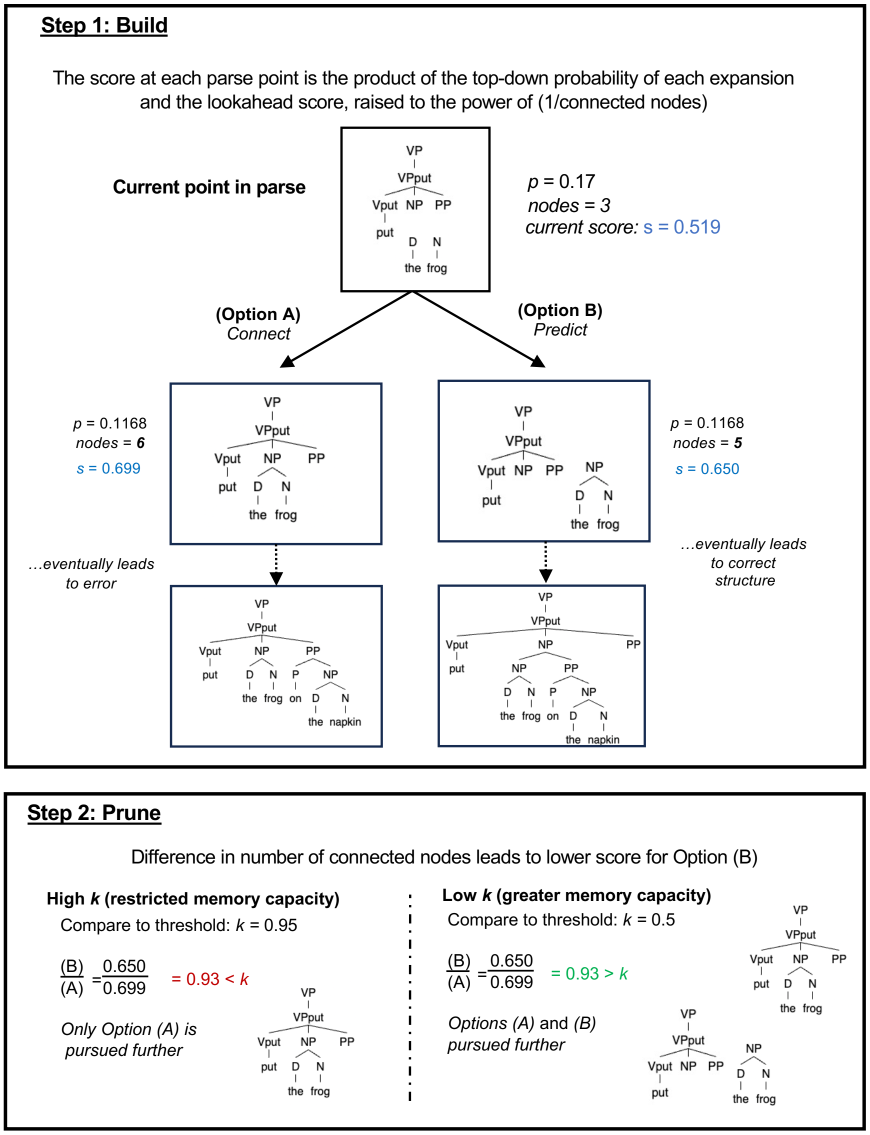


Figure 6.2: Partial analyses corresponding to the kindergarten path and intended analysis.

A working memory restriction could be the culprit here: if working memory is constrained to a sufficient degree, the intended analysis could fall off of the beam and lead the parser to pursue the garden path. We demonstrate this possibility with a toy example in Figure 6.3, where a failure to find the intended parse could arise from the crucial choice point where the parser is scoring two candidates: one in which the parser CONNECTs to the *NP* predicted from *VPput*, completing the predicted *NP*, and one where the parser PREDICTs an *NP* on the basis of the corner *D, N*, without connecting it to the main structure, thereby allowing it to eventually be the left corner of a larger *NP* with an adjunct.

It is the analysis resulting from the PREDICT which corresponds to the intended structure for the sentence *Put the frog on the napkin in the box.* The crucial difference between these two candidates is the number of resulting connected nodes. In the case of a CONNECT, the number



Step 1: Build

The score at each parse point is the product of the top-down probability of each expansion and the lookahead score, raised to the power of (1/connected nodes)

Step 2: Prune

Difference in number of connected nodes leads to lower score for Option (B)

Figure 6.3: Parsing with varying k values, showcasing the crucial choice point.

of connected nodes is 6, while in the case of a PREDICT, the number of connected nodes is 5. Because the parser favors more connected nodes, at high enough values of k, the candidate which leads to the intended structure is pruned, resulting in the inability to pursue this analysis on the basis of later-arriving input.

Thus, in order to model a progression in working memory capacity over development, higher values of k correspond to a more restricted working memory: only those parses with the highest scores can be maintained, while any candidates with scores that are at a ratio below the threshold

k (when compared to the highest score) are discarded. In a mature parser with fully developed working memory, however, k may be a lower value. By virtue of this lower threshold, more candidates can be maintained on the beam at any given point, thus allowing the correct analysis to be pursued at a later point if a formerly-pursued analysis becomes incompatible with the input. This fact provides a natural parallel to the exact findings we wish to simulate: while adults are able to maintain multiple candidate analyses and revise their initial parses even if the initial parse was more distributionally favored, children tend to only pursue the most distributionally favored candidates and systematically overlook the rest, thus prohibiting revision. It is unlikely, however, that $k = 0$ in the mature parser, however: the beam must be subject to some baseline memory constraint because memory is a finite cognitive resource.

6.3.2 Cognitive control

Similar to Ness et al. (2022) and Ovans (2022), we adopt a model in which cognitive control is responsible for "suppressing" partial analyses that are not relevant for the particular task at hand: parsing a particular string. We implement this with a weighting parameter, which, if non-zero, causes the score of the partial analysis to be unduly influenced by the score of its parent in the search tree. Intuitively, if cognitive control is fully engaged, the weight on the score of the parent of a partial analysis should be 0. In other words, any new steps should be considered solely on the basis of their fit with the new bottom-up information or top-down structure being considered at that point, independent of whether the parent of the partial analysis was high-scoring before this new information was considered. Thus, when cognitive control is fully engaged, any newly-arriving bottom-up information that is incompatible with a given candidate should immediately indicate that this particular candidate will not yield a suitable analysis, giving a score of 0. For example, in an adult-like system, upon encountering *in* in the sentence *Put the frog on the napkin in the box*, the VP-attached partial analysis in which *on the napkin* is the destination for the frog should be suppressed (i.e., given a score of 0) since it can no longer contribute to a single connected parse for the string. On the other hand, if cognitive control is not fully engaged (or still developing), a lag in score updating may occur: because the parent of a partial analysis was a very high-scoring partial analysis, it retains some of this score. Thus, even if its score should be lower given the newly-considered information, the score of this partial analysis will remain artificially high because its previously-high score is being factored in. In our test sentence, *Put the frog on the napkin in the box*, this would mean that if the VP-attached partial analysis was initially very high-scoring, it would retain some amount of this high score, allowing the erroneous candidate to remain in competition with other partial analyses.

We formalize this as a weighted geometric mean. As the parser considers the possible next steps, the score of a new analysis, s_{super}, is calculated through the weighted geometric mean of the score of the current step and the step prior. The formula for this calculation is given in (39), where *w* is the weight of the score on the parent in the search tree, s_{prev} is the score of the parent in the search tree, and s_{step} is the score of the newly-formed partial analysis, both prior to any weighting being applied and calculated as described in §6.2.2:

(39) $$s_{super} = \sqrt[w+1]{(s_{prev})^{w} \times s_{step}}$$

Thus, if $w = 0$, s_{super} would simply equal s_{step}. If $w > 0$, on the other hand, s_{super} would factor in the previous score of the candidate. This models a particular way in which immature cognitive control could interact with parsing by causing a delay in score updating: in an immature system, the ability to update the candidate's score on the basis of novel information is delayed, thereby modeling difficulty with rapidly incorporating novel input (Ness et al., 2022; Ovans, 2022; Trueswell et al., 1999; Woodard et al., 2015).

In the simulations that follow, we test the contributions of cognitive control limitations by manipulating the value of *w* while holding *k* fixed to a small value (specifically, 0.05), which represents the baseline degree of working memory constraints present even in the mature parser. We hold *k* constant to allow us to isolate the role of developing cognitive control in parsing the test sentences from Trueswell et al. (1999). At sufficient values of *w*, the candidate that would lead to the intended analysis would be "pushed off the beam" by other analyses that are initially high-scoring and whose scores haven't been updated quickly, leading to a score discrepancy exceeding k. Thus, in effect, the cognitive control parameter w controls how working memory is allocated within the beam (Baddeley & Hitch, 1974; Botvinick et al., 2001).

6.4 Simulation 1

We evaluate the contribution of each of the model's parameters to performance with 'kindergarten path' sentences from Trueswell et al. (1999). In particular, 'success' entails deriving the reported asymmetries with the temporarily-ambiguous and unambiguous sentence, repeated in (38).

(37) a. Put the frog on the napkin in the box (Temporarily-ambiguous)

b. Put the frog *that is* on the napkin in the box (Unambiguous)

With child-like settings (i.e., either a high weighting on the previous score of the analysis, or a more restricted beam), we expect the parser to fail to find the intended structure corresponding to the temporarily-ambiguous sentence, namely a structure where the PP *on the napkin* is an adjunct to the NP *the frog*. At these same settings, the parser should successfully find the appropriate structure corresponding to the unambiguous sentence. With adult-like settings of each parameter (e.g., either a low weighting on the previous score of the candidate, or a less restricted beam), the parser should find the intended structure for both the ambiguous and unambiguous sentence. A visualization of success in the parser can be found in Table 6.2.

Here, a check mark indicates that the parser has returned the intended structure for the sentence (either ambiguous or unambiguous), while a dash indicates that it has not. As discussed in the previous section, smaller settings of the relevant parameter reflect a more mature system: a smaller k reflects a larger beam, while a smaller w indicates less weighting on the previous scores.

As discussed in Chapter 5, the literature has proposed two hypotheses for children's performance with temporarily-ambiguous sentences like *Put the frog on the napkin in the box.* On one

	Temporarily-ambiguous	*Unambiguous*
	✓	✓
More mature	✓	✓
	✓	✓
	✓	✓
	✓	✓
	-	✓
Less mature	-	✓
	-	✓
	-	✓

Table 6.2: Successful modeling of the developing parser with respect to Trueswell et al. (1999).

hypothesis, limited working memory capacity, but not limited cognitive control, is responsible for children's seeming inability to interpret these sentences (Montgomery et al., 2008; Trueswell et al., 1999; Weighall & Altmann, 2011; Zhou et al., 2021). We test this hypothesis by varying k incrementally while holding w constant at 0, modeling a cognitive control system that interacts with parsing in an adult-like way. Note that we do not argue that $w = 0$ corresponds to a fully mature cognitive control system: we know from the literature on developing cognitive control that this system does not reach full maturity until early adulthood (see Chapter 5, §5.3.2.1). Rather, setting w to 0 reflects the strong hypothesis that whatever development is occurring in children's cognitive control systems, it does not affect their ability to update parse scores (see Ness et al., 2022; Ovans, 2022). Thus, if there exists some value(s) of k when w=0 such that the parser finds the intended structure for the unambiguous sentence, *Put the frog that is on the napkin in the box*, but not for the temporarily-ambiguous sentence, this would suggest that a limited working memory capacity alone could be responsible for children's performance.

On the other hypothesis, an immature cognitive control system is responsible for children's performance with the temporarily-ambiguous sentence (Hsu et al., 2021; Novick et al., 2005; Ovans, 2022; Woodard et al., 2016). We test this hypothesis by varying w incrementally and holding k constant at a baseline value of 0.05. As with the previous hypothesis, we do not presume that this baseline value necessarily reflects a fully adult-like working memory capacity; instead, we posit that this reflects an adult-like *interaction* between working memory and parsing, namely an adult-like degree of working memory limitations on sentence processing. If there exists some value(s) of w when $k = 0.05$ such that the parser finds the intended structure for the unambiguous sentence but not the temporarily-ambiguous sentence, this would suggest that the developing ability of immature cognitive control to allocate the baseline amount of working memory could be responsible for children's performance. Lastly, we will also be testing additional combinations of k and w to examine the parser's performance when both systems are immature.

Rule	*p*	*Rule*	*p*	*Rule*	*p*
ROOT → FRAG	0.220	CP → C VPis	0.006	N → box	0.0040
ROOT → FRAG FRAG	0.002	CP → C XP	0.994	N → frog	0.0005
ROOT → XP	0.778	C → that	0.025	N → other	0.9904
FRAG → VP	0.107	C → other	0.975	Vis → is	0.700
FRAG → PP	0.050	**PP → P NP′**	**0.827**	Vis → other	0.300
FRAG → XP	0.843	PP → XP	0.173	**NP′ → D N**	**1.000**
VP → VPput	0.010	Vput → put	1.000	NP → D N	0.123
VP → XP	0.990	D → the	0.450	NP → NP PP	0.013
VPput → Vput NP PP	0.300	D → a	0.310	NP → NP SBAR	0.008
VPput → Vput XP	0.700	D → other	0.240	NP → XP	0.856
VPis → Vis PP	0.037	P → on	0.140	SBAR → C S	0.224
VPis → Vis XP	0.963	P → in	0.180	SBAR → XP	0.776
XP → other	1.000	N → napkin	0.0007	S → VPis	0.259

Table 6.3: Partially-lexicalized PCFG based on child-directed speech

In Simulation 1, we investigated our model's performance with the temporarily-ambiguous and unambiguous sentences in Trueswell et al. (1999). The parser used the PCFG shown in Table 6.3. The same grammar was used in Simulations 1 and 2, with the exception of two rules (printed in bold face) which were only present in Simulation 1. These rules were included in Simulation 1 to enforce a somewhat-artificial requirement that NP complements of P do not have adjuncts, thereby disallowing an alternative analysis of the temporarily-ambiguous sentence that was not considered in Trueswell et al. (1999). By disallowing this alternative analysis, we are able to compare only the analyses that have been considered experimentally; we return to this alternative analysis in Simulation 2.

The PCFG was developed by extracting the rules needed to parse the sentences used in the simulations that follow (i.e., the sentences in Trueswell et al., 1999) from the CHILDES Treebank

(Pearl & Sprouse, 2013). We used a custom-written script to manipulate the trees to include the lexical information of each verb at the *V* and *VP* level. The weights in the PCFG were estimated by counting the relative frequencies of each expansion of a given nonterminal in the parse trees of the Treebank. The resulting PCFG was then checked for any issues. Based on the conventions of the CHILDES Treebank (Pearl & Sprouse, 2013), all of the imperatives in Trueswell et al. (1999) are generated as *VP* fragments. The grammar also allows for an utterance to consist of two otherwise disconnected fragments via the *ROOT* → *FRAG FRAG* rule, which is needed to model the garden-path analysis of the temporarily-ambiguous sentence in Trueswell et al. (1999) because this analysis consists of two fragments: a *VPput* (*Put the frog on the napkin*) and a *PP* (*on the napkin*; see Figure 6.4, left). Lastly, *XP* is used to capture the distribution of all other rules in the TreeBank, allowing us to accurately reflect the probabilities of the rules necessary to parse the test sentences while abstracting away from unneeded rules. *XP* can only ever rewrite to *other*, which is never present in the input string for the parser.

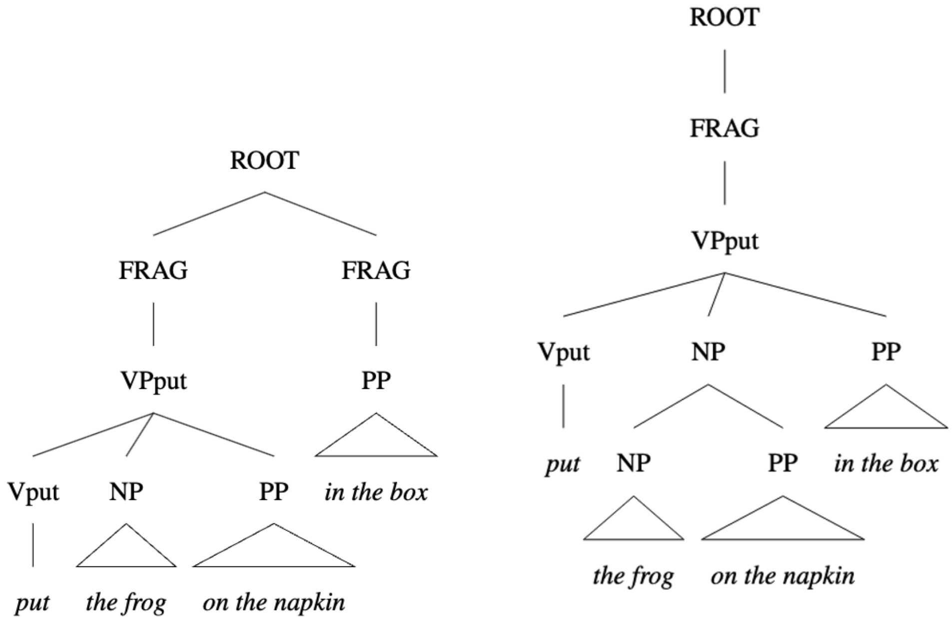


Figure 6.4: The analysis analogous to being "kindergarten-pathed", left, and the intended interpretation, right.

6.4.1 Results

Working memory (k): More restricted

Working memory (k)	0	1	2	3	4	5	6	7	8	9	10
0.275	-	-	-	-	-	-	-	-	-	-	-
0.25	-	-	-	-	-	-	-	-	-	-	-
0.225	✓	-	-	-	-	-	-	-	-	-	-
0.2	✓	-	-	-	-	-	-	-	-	-	-
0.175	✓	-	-	-	-	-	-	-	-	-	-
0.15	✓	-	-	-	-	-	-	-	-	-	-
0.125	✓	✓	-	-	-	-	-	-	-	-	-
0.1	✓	✓	✓	-	-	-	-	-	-	-	-
0.075	✓	✓	✓	✓	-	-	-	-	-	-	-
0.05	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-

Working memory (k): Less restricted

More updating ← Cognitive control (w) → *Less updating*

Table 6.4: Model performance on the temporarily-ambiguous sentence. Check-marks denote intended parses of the sentence, while dashes denote failures to find the intended parse. The intended parse for the unambiguous sentence is found at all of the parameter settings shown.

The results for Simulation 1 can be found in Table 6.4. Holding the cognitive control parameter constant at $w = 0$ and varying the working memory parameter k, we find that the parser can successfully find the intended structure for the unambiguous sentence *Put the frog that is on the napkin in the box* for all values of k. By contrast, the parser fails to find the intended structure for the temporarily-ambiguous test sentence in Trueswell et al. (1999) at more restrictive (higher) values of k, above 0.225. When the beam is more permissive at smaller values of k – analogous to an adult-like interaction of working memory with parsing – the parser is able to find the intended structure for both the temporarily-ambiguous sentence and for the unambiguous sentence. Meanwhile at more restrictive values – analogous to smaller working memory capacity, as in children – the performance of the parser mirrors that of a 5-6 year old child in Trueswell et al. (1999), unable to find the intended parse of the temporarily ambiguous sentence, but able to find the intended parse for the unambiguous sentence.

Conversely, when holding the working memory parameter k constant at $k = 0.05$, and varying the inhibitory control parameter, w, the parser can also successfully capture the asymmetry observed in Trueswell et al. (1999). Namely, at lower values of w (which reflect more rapid score updating for candidates and thus an adult-like interaction of cognitive control and parsing) the parser successfully finds the intended structures for both the ambiguous and unambiguous test sentence. However, at higher values of w (reflecting a lag in updating a candidate's score), the parser is no longer able to find the correct structure for the temporarily-ambiguous sentence.

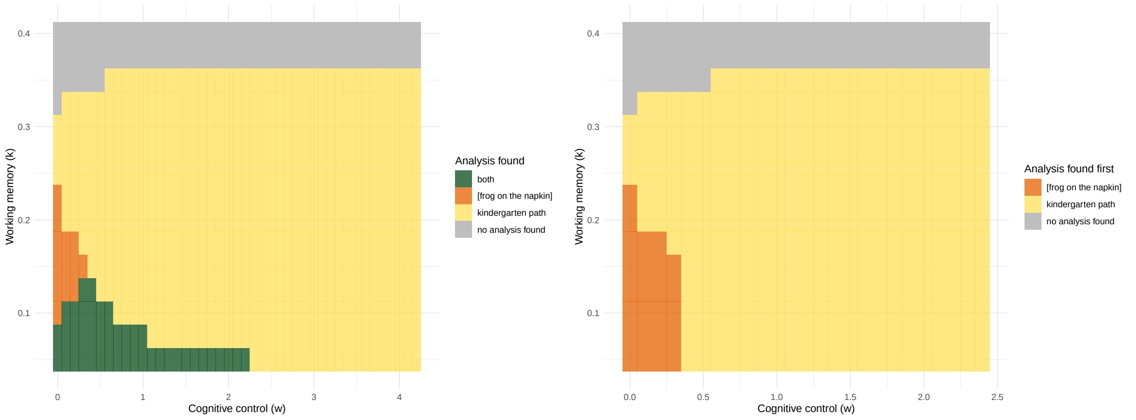


Figure 6.5: Analyses found for the temporarily-ambiguous sentence, left, and distribution of the analysis found first, right. The bottom left corner reflects greater maturity in both systems.

We further examined what caused the intended analysis not to be found. In particular, we investigated whether the kindergarten-path analysis, with disconnected fragments (see Figure 6.4) is out-competing the candidate that would result in the intended, fully-connected analysis. This structure, in which *on the napkin* is parsed as the goal, could potentially out-compete the candidate that would result in the intended structure (in which *on the napkin* is the PP adjunct sister to the NP *the frog*; shown again in Figure 6.4, right) at some point during parsing. To probe this possibility, we identified combinations of k and w at which the parser finds a "kindergarten path" structure (Figure 6.4, left), and compared its distribution to cases in which the intended structure was found (Figure 6.5). In these figures, the yellow regions indicate the combinations of k and w at which this "kindergarten path" analysis was found. In Figure 6.5, left, we depict all of the analyses found at every combination of k and w, with dark green representing both analyses being found, orange representing only the intended structure being found, and grey representing no analysis found. In regions where both analyses were found, the intended analysis is always higher-scoring. However, although the intended analysis is always the highest scoring, it is not always found first, as shown in Figure 6.5, right. Here, we see that the intended analysis is only found first for certain values of k ranging from 0.05 to 0.225, when w is between 0 and 0.3. At all other combinations of k and w where any analysis is found, the kindergarten path analysis is found first.

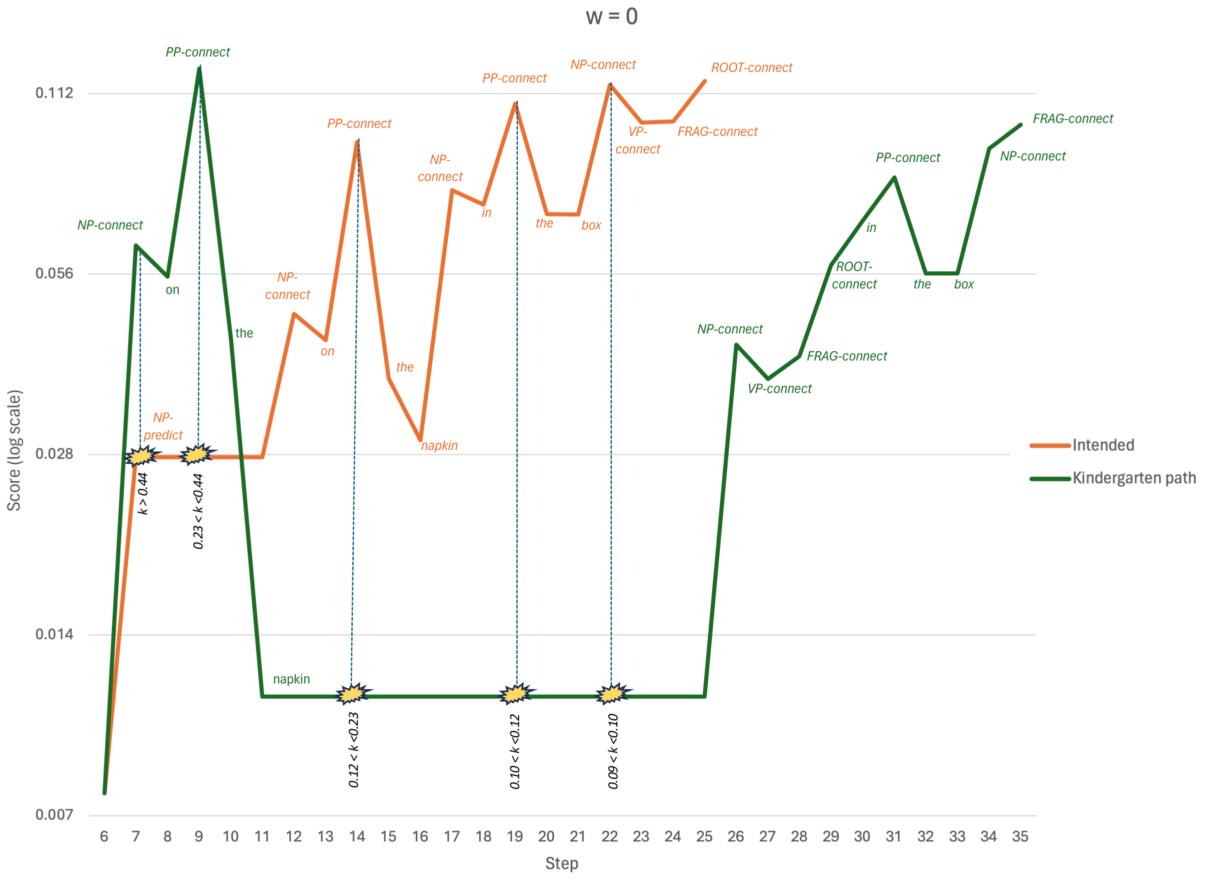


Figure 6.6: Competition between the "kindergarten path" analysis and the intended analysis when $w = 0$. "Crash" points, where the lower scoring analysis could be lost, are indicated with a yellow symbol. The lower scoring analysis will be lost at the specified points where k exceeds the ratio of the scores at that point. For example, for any k greater than 0.44, the intended candidate would be lost once the kindergarten path candidate CONNECTs *the frog* to the NP in Step 7. Similarly, for any k greater than 0.23, the intended candidate would be lost once the kindergarten path analysis CONNECTs *on* to the PP in Step 9.

To identify the point in parsing at which the kindergarten path analysis out-competes the intended analysis as a function of each developing system, we plotted the scores of each partial analysis for each step of parsing. We first begin with a case where $w = 0$, simulating a mature interaction of cognitive control with parsing. Figure 6.6 shows the competition between the two partial analyses as they are pursued. The difference between connecting the NP to the VPput structure and predicting an NP, which will later serve as the left corner of an NP with a PP adjunct, is responsible for the initially lower score of the intended analysis compared to the kindergarten path analysis. Here, the y-axis shows the score of each candidate at a given parsing step; a candidate is 'lost' if the ratio of that candidate's score and the score of the highest-ranked analysis drops below k. We see that the crucial choice point of how to integrate *the frog* (as shown in Figure 6.3) ultimately causes the intended analysis to be lost for any k greater than 0.23 (Step 9 in the parse). This is due to the fact that the kindergarten-path partial analysis is pursued first, and bottom-up information continues to support this candidate for a time until it eventually reaches a score that is 4.35 times greater than that of the intended analysis (in other words, the score of the intended partial analysis is only 0.23 that of the kindergarten path analysis). When k is more restrictive than this ratio, the intended candidate will be dropped from the beam because it did not meet the k threshold.

On the other hand, when w is high, the intended analysis cannot survive even at the least restrictive values of k. Figure 6.7 shows that when $w = 8$, even at very low values of k (e.g., $k = 0.05$), the intended candidate cannot survive past the point that kindergarten path analysis SHIFTs *the* (Step 10). At this point, the kindergarten path analysis reaches a score that is 25 times greater than that of the intended analysis (in other words, the intended analysis has a score that is only 0.04 of the kindergarten path analysis). As a result, even with minimal restrictions on memory ($k = 0.05$), the intended analysis does not meet this threshold and is dropped from the beam. Note that although $k = 0.05$ represents a restriction on working memory, it is otherwise permissive: when $k = 0.05$ and w is low, the intended parse is found. Thus, although working memory is always restricted in this parser, it is the way that a high w allocates memory within the restricted beam that causes the intended parse to be lost. This suggests that an immature interaction of cognitive control with parsing can lead to failure to find the intended parse, even with a large capacity on the

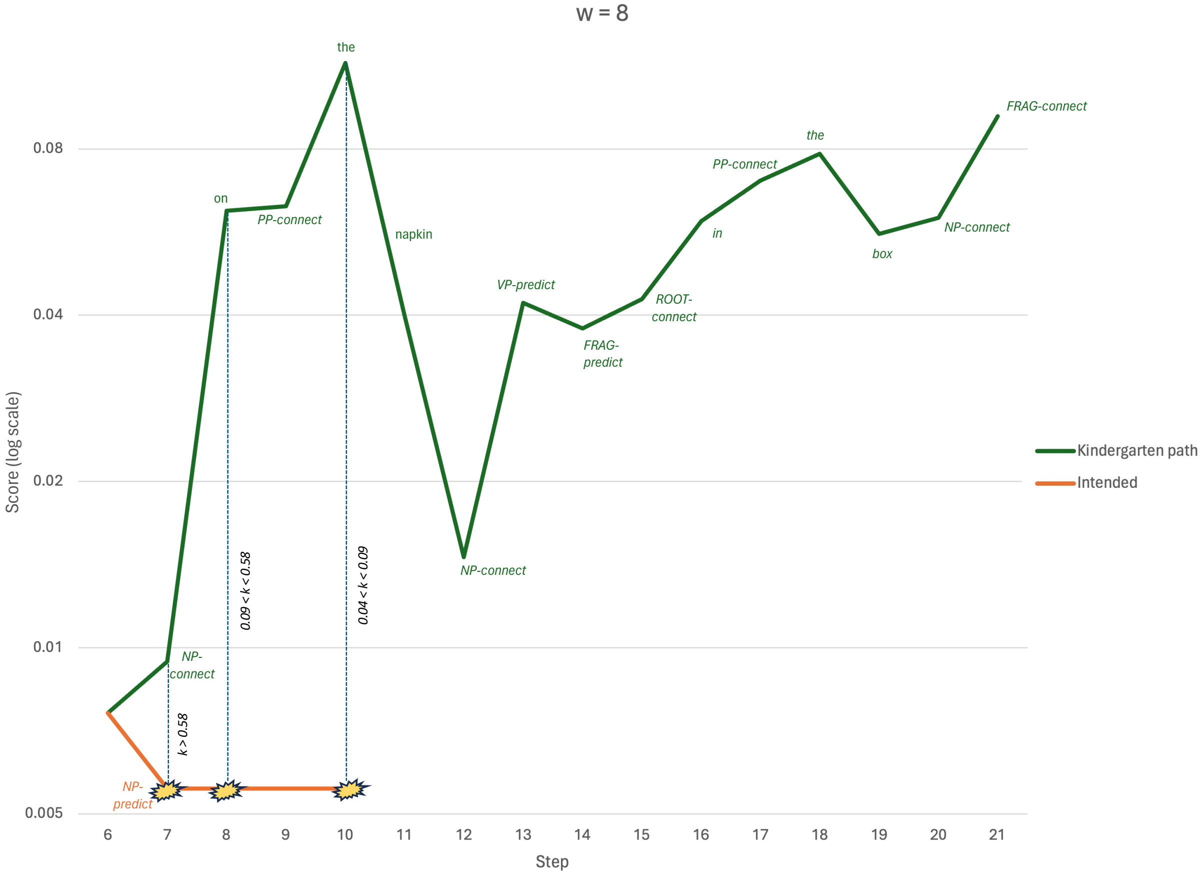


Figure 6.7: Competition between the "kindergarten path" analysis and the intended analysis when w = 8. The ratio of scores, indicating the minimum k value at which the lower-scoring analysis would survive, is denoted in the droplines. At each step, if k is smaller than the indicated value, the lower-scoring analysis would be dropped from the beam.

beam.

6.4.2 Discussion

The findings of Simulation 1 suggest that either restrictions on working memory capacity, immature cognitive control, or a combination of both, could be responsible for the difference between children and adults' performance with temporarily-ambiguous sentences like *Put the frog on the napkin in the box*. Namely, varying either of our two parameters, while holding the other fixed, can account for the findings in Trueswell et al. (1999): there are some values where the parser fails to find the intended analysis and finds only the kindergarten-path analysis for the temporarily-ambiguous sentence, and at those same values, the parser always finds the intended analysis for the unambiguous sentence. Our model is consistent with a wealth of experimental work that has found correlations between the efficiency of sentence processing in adults and children and measures of working memory and cognitive control (Caplan & Waters, 1999; Hsu et al., 2021; Hsu & Novick, 2016; January et al., 2009; Lewis, 1996; Montgomery et al., 2008; Novick et al., 2014; Novick et al., 2009; Novick et al., 2005; O'Rourke, 2013; Weighall & Altmann, 2011; Woodard et al., 2016; Ye & Zhou, 2009; Zhou et al., 2021). However, this simulation does not identify whether development in working memory alone, or development in cognitive control, is primarily responsible for this effect.

Our investigation of the distribution of the analyses found across different values of *k* and *w* (Figure 6.5) highlighted that the "kindergarten path" analysis is widely available: the parser finds a "kindergarten path" analysis at a large range of *k* and *w*'s. At most of these combinations of *k* and *w*, the parser can only find this analysis; for a small subset of the combinations, it finds both the kindergarten path analysis and the intended analysis. This suggests that when the working memory and/or cognitive control systems are less mature (at greater values of *k* and *w*), the system is "kindergarten path"-ed by this initially-promising structure. However, as the system(s) responsible mature such that they interact with parsing in a more adult-like way, the parser is also able to find the intended analysis, which is always higher-scoring. With respect to which analysis is found first when both analyses are available, the intended analysis is only found first at certain combinations of *k* between 0.05 and 0.225 and *w* between 0 and 0.3. At all other combinations of *k* and *w* where both analyses are found, the kindergarten path analysis is found first. Cases in which the kindergarten path analysis is found first could potentially be viewed as a mild form of garden-pathing, where children initially pursue this structure, but are still able to backtrack to pursue the candidate corresponding to the intended analysis because it hasn't been lost from the beam.

The results in Simulation 1 are consistent with either immature working memory or cognitive control, or a combination of both, being primarily responsible for children's performance on experimental tasks. In other words, these two possibilities do not make different predictions when considering the intended analysis for this sentence. However, as we alluded to earlier, there is an additional analysis of the sentence available where these two possibilities may make different predictions. Recall that in this simulation, the grammar that was used explicitly disallowed *NP* complements of *P* from having adjuncts, in order to rule out this alternative analysis and restrict the simulation to only the two analyses that had been previously considered experimentally. This was done for the sake of convenience, and was not intended to be a meaningful claim about the rules that children have for generating *PP*s.

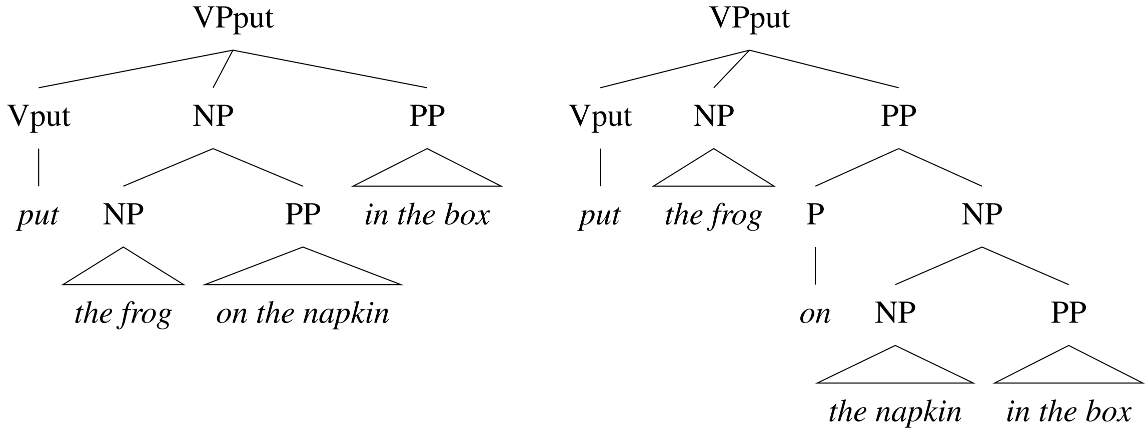


Figure 6.8: Global ambiguity in Trueswell et al. (1999)'s temporarily ambiguous test sentence. The structure on the left (*on*-as-modifier) corresponds to the interpretation compatible with the layout in the study; the structure on the right (*on*-as-goal) is incompatible.

Under the more realistic assumption that *NP* complements of *P* can actually have adjuncts, the test sentence *Put the frog on the napkin in the box* can also be analyzed under the structure in Figure 6.8, right. Here, *the napkin in the box* is interpreted as the destination for *the frog*. Crucially, this interpretation is not made available given the scene in the Trueswell et al. (1999) studies, because there is no 'napkin in the box' in the scene. Although the grammar in this simulation was likely not realistic, it served the purpose of allowing us to model a parser that could use this top-down scene information to avoid ever considering the contextually-unsupported analysis. However, we know from the literature on early sentence processing that children do not appear to easily incorporate visual or referential context, relying instead on verb biases from their knowledge of argument structure (Lidz et al., 2017; Omaki et al., 2014; Snedeker & Trueswell, 2004; Trueswell et al., 1999). So, although children seem unable to find the intended analysis in these previous experiments, is it possible that they might still find this alternative, albeit contextually-unsupported, analysis? This may be plausible: given the more realistic assumption that the *NP* complement of *P* can have an adjunct, the undesired partial analysis in Figure 6.3 (where the parser makes a CONNECT rather than a PREDICT early on, making 'the frog' the entire theme NP) can actually lead to a complete analysis that is compatible with the sentence (Figure 6.8, right). We refer to this alternative analysis as the *on*-as-goal analysis.

If children or adults were able to find this parse, despite having no contextual support, they would have no way of demonstrating it in the context of these experiments. Thus, if children's difficulty in parsing *Put the frog on the napkin in the box* in Trueswell et al. (1999) arises from the parser 'losing' the intended analysis, which PREDICTs an NP rather than immediately CONNECTing it (as suggested by Simulation 1), we might still expect them to be able to find the *on*-as-goal analysis even with restricted working memory or cognitive control. We test this prediction in a second simulation.

6.5 Simulation 2

In this simulation, we examine the parser's ability to find the alternative *on*-as-goal analysis, and test how cognitive control and working memory contribute to this ability. As in Simulation 1, we do this by incrementally varying k while holding w constant, and vice versa. In Simulation 2, we use the PCFG in Table 6.3 with the boldface rules removed and a $PP \rightarrow P\ NP$ rule added. Recall that the PCFG used in Simulation 1 included a NP$'$ nonterminal. This nonterminal was intentionally included to rule out the additional *on*-as-goal analysis ([napkin in the box]) by barring the goal *PP* from having a complex *NP* daughter containing another *PP*. In doing so, we could examine how the intended analysis competes against the kindergarten path analysis. By removing this artificial restriction on the grammar, the parser will be able to find the *on*-as-goal analysis shown in Figure 6.8, where *on the napkin in the box* is interpreted as the goal. This allows us to see how the *on*-as-goal analysis competes against both the kindergarten path and the originally-intended analysis ([frog on the napkin]). As a result, Simulation 2 models a parser that does not use top-down scene information to constrain the analyses it considers.

6.5.1 Results

The analyses found at various combinations of k and w are shown in Figure 6.9. We find that, for almost all combinations of w and k in the range tested, the *on*-as-goal analysis is found. When k is fixed at a low value, varying w from big to small has the following effect: at large w, the parser finds both the *on*-as-goal analysis and the kindergarten path analysis. As w decreases, the parser finds all of the possible analyses, including the intended *on*-as-modifier analysis. However, when w is fixed at 0, varying k from big to small has a different effect. At high values of k, the parser initially finds only the *on*-as-goal analysis. As k decreases, the kindergarten path is also found. However, at k's between 0.175 and 0.300, the parser only finds the originally intended, *on*-as-modifier analysis. Finally, as k decreases beyond 0.175, the parser finds all three analyses. At no combination of k and w does the parser return the kindergarten path analysis as the only analysis found: at all points where the kindergarten path analysis is found, the *on*-as-goal analysis is also found.

As in Simulation 1, we also identified which analysis is found first by the parser at each combination of k and w (Figure 6.9, right). We find that the *on*-as-goal analysis is found first at all combinations of k and w tested, with the exception of when $w = 0$ and $k \leq 0.300$, and $w = 0.1$ and $k \leq 0.25$, where the intended analysis is found first. In part of this range (where $0.2 \leq k \leq 0.300$ when $w = 0$ or $0.227 \leq k \leq 0.275$ when $w = 0.1$) the intended analysis is the only parse found.

Importantly, now that we are considering the sentence to be globally as well as locally ambiguous, we find a different distribution of complete analyses found along the x- and y-axes of this

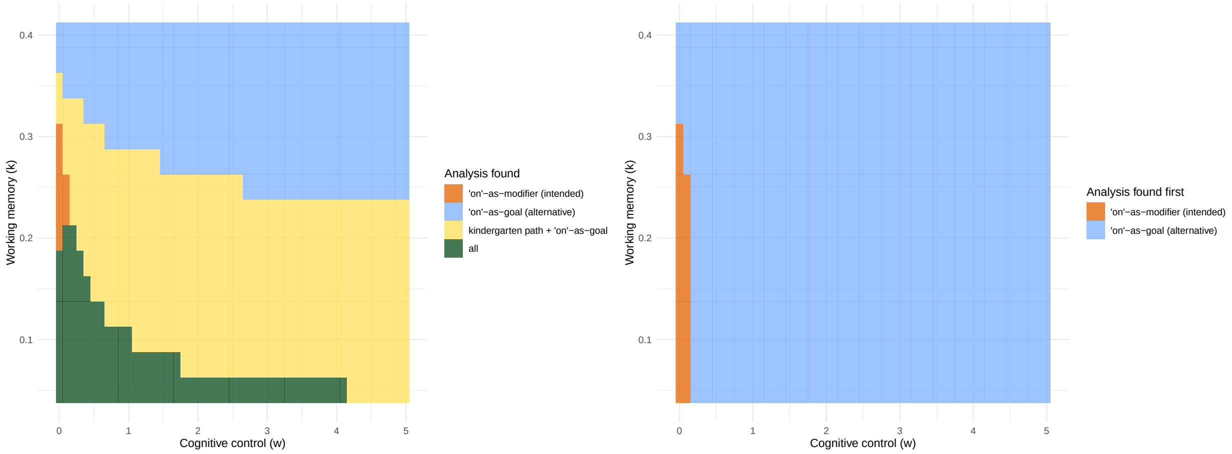


Figure 6.9: Cognitive control and working memory constraints make different predictions for the temporarily-ambiguous sentence with respect to analyses found.

graph. Holding w constant at 0 and varying k, we find a region where the alternative *on*-as-goal analysis fails to be found (orange region in Figure 6.9, left). There is no such region found when holding k constant at 0.05 and varying w.

How does only the intended analysis survive in this region (e.g., the orange region in Figure 6.9, left)? In cases where only the intended analysis survives, the *on*-as-goal candidate is ultimately lost due to a cross-over in scores between it and the intended analysis. Because the candidate scores initially diverge at the crucial choice point between CONNECTing and PREDICTing the NP *the frog* (see Step 7 in Figure 6.3), the candidate which can yield the *on*-as-goal analysis is initially pursued. The intended analysis remains stagnant throughout several parse steps as this alternative is expanded. However, once *napkin* is SHIFTed in Step 11, the score of the *on*-as-goal candidate dips below that of the intended analysis. The intended analysis is subsequently pursued, ultimately reaching high scores that "kill off" the alternative analysis. In the region where both analyses are found but the intended analysis is found first, the same "cross-over" is observed; however, because this region is at lower k's ($k \leq 0.19$), the *on*-as-goal analysis is never lost and can still be pursued once the intended analysis is found.

In cases where only the alternative *on*-as-goal analysis and kindergarten path partial analysis survive, the initial divergence continues to grow until the ratio of the intended partial analysis to

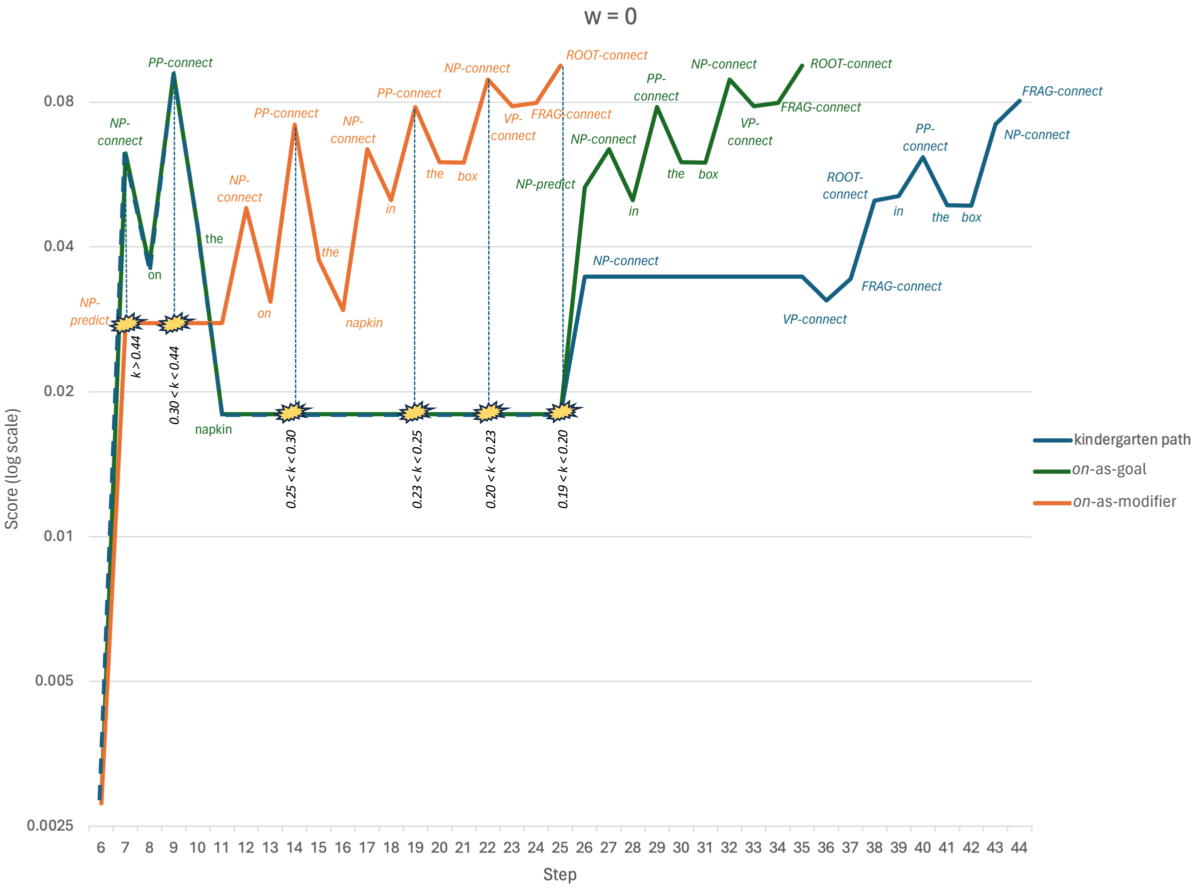


Figure 6.10: Competition between the intended analysis and the *on*-as-goal analysis when w = 0 across parser steps. As before, yellow symbols indicate "crash" points for the lower scoring analysis, where if k is at or above the ratio indicated, that analysis will be lost.

the *on*-as-goal partial analysis is lost. For example, Figure 6.11 depicts the competition between the alternative *on*-as-goal analysis and the intended analysis at a high w. Just as in Simulation 1, the alternative *on*-as-goal candidate is pursued first due to the initially-higher score at the crucial choice point. Eventually, this candidate achieves such a high score that the intended candidate is lost, because the ratio of the scores is 0.04 at the point that the VPput structure is CONNECTed to the VP in Step 19. Thus, even when the beam is only slightly restricted (e.g., k = 0.05), the intended analysis is lost at this point and cannot be pursued further.

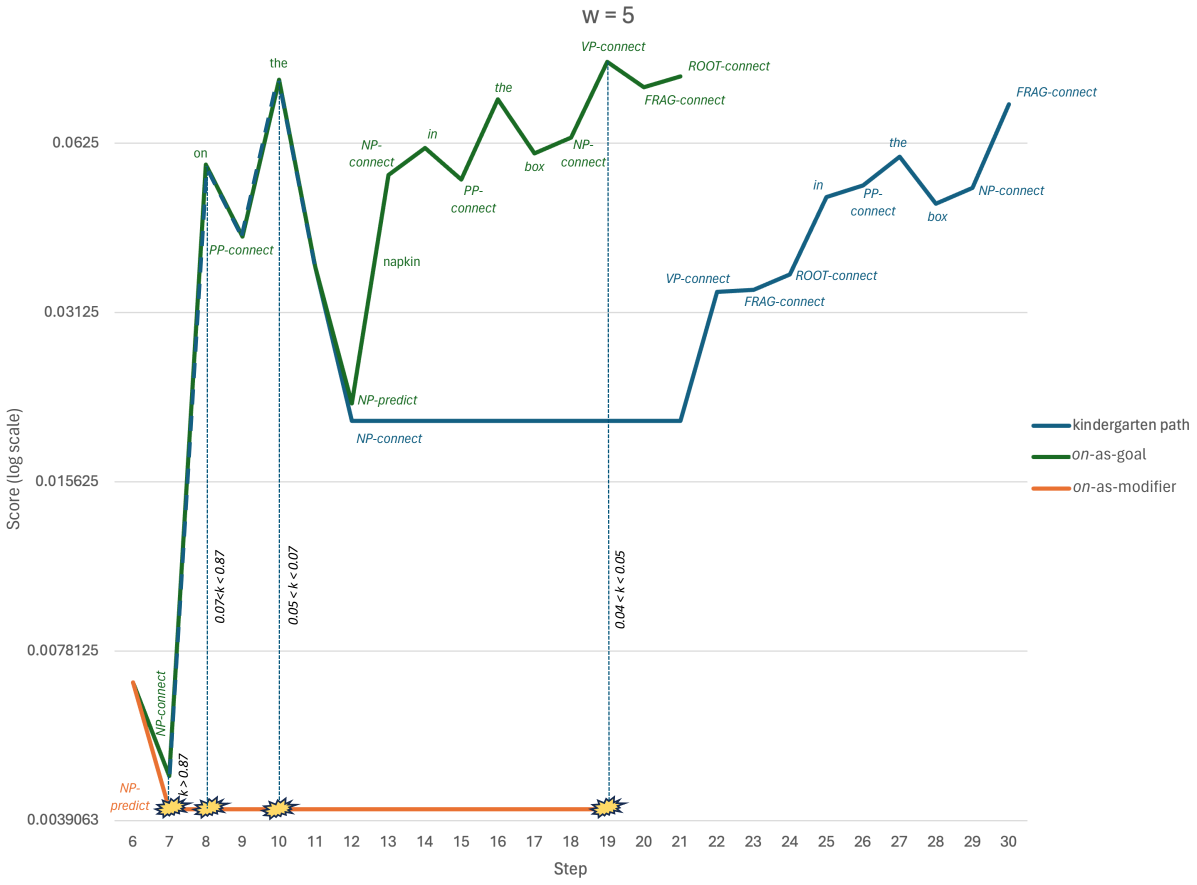


Figure 6.11: Competition between the intended analysis and the *on*-as-goal analysis when k = 0.05 and w = 5 across parser iterations. Text denotes relevant transitions and yellow symbols indicate possible "crash" points.

6.5.2 Discussion

The findings from Simulation 2 suggest that children's performance with the test sentence in Trueswell et al. (1999) could be the result of not merely competition with a kindergarten-path structure, but also an alternative analysis of the sentence: one in which *the napkin in the box* is the destination for the frog. Although this analysis is incompatible with the visual scene in the original experimental set-up, children's insensitivity to visual context when parsing (Snedeker & Trueswell, 2004; Trueswell et al., 1999) could cause this analysis to remain in competition with the intended analysis, potentially causing the partial intended analysis to be lost if working memory or cognitive control are still developing and impose stricter limitations on the candidates pursued.

Recall that Simulation 1 offered no way of differentiating developing working memory and cognitive control systems as primarily responsible for children's parsing differences. In Simulation 2, however, our parser makes different predictions for these two possibilities. Specifically, Simulation 2 makes empirical predictions with respect to the developmental time course of when children could find the alternative *on*-as-goal ([napkin in the box]) interpretation as compared to the interpretation tested in previous experiments, *on*-as-modifier ([frog on the napkin]). Varying *w* while holding *k* constant at 0.05 (or other values), the parser finds the *on*-as-goal analysis but not the *on*-as-modifier analysis at larger values of *w*, where score updating is slower. As *w* decreases, modelling increased score-updating capabilities, the parser finds both analyses. Thus, under the hypothesis that an immature cognitive control system impacts sentence processing and is responsible for children's seeming inability to parse the test sentence in Trueswell et al. (1999), our model predicts that that there is a point in development where children can only find the alternative analysis; as cognitive control matures, they will find both.

Conversely, varying *k* and holding *w* constant at zero, the parser finds only the *on*-as-goal analysis at higher values of *k*. As *k* decreases, the parser can only find the *on*-as-modifier analysis. Finally, as *k* decreases further and approaches 0, the parser can find both analyses. Under the hypothesis that limited working memory capacity, but not limited cognitive control, is responsible for children's performance, our parser predicts that early in development, only the alternative analysis can be found, followed by a period when only the previously-tested interpretation can be found, and then finally both. Notice that there was a similar region in Simulation 1 (Figure 6.5), where only the intended analysis and not the kindergarten-path analysis was found. But the presence of this region in Simulation 1 does not make any predictions that would be observable in children's behavior with these sentences: children would simply shift from being kindergarten-pathed to being able to find the intended analysis, regardless of whether *k* or *w* is being varied. On the other hand, the presence of this region in Simulation 2 makes the prediction that children will be able to find qualitatively different complete analyses at different stages of development, depending on which parameter is changing.

Thus, when looking at an analysis that has not previously been considered, the two hypotheses about which cognitive system is primarily responsible for children's parsing difficulties make different predictions about their developmental time course as each system matures. These predictions could be tested empirically in a design that reveals whether children can, in fact, find this alternative analysis. We propose such a design in the general discussion.

6.6 Summary and general discussion

In this chapter, we implemented a formal model of predictive, incremental parsing in development, with parameters corresponding to working memory and cognitive control. With this model, we were able to capture the empirical findings of the seminal Trueswell et al. (1999) paper, finding states of developing working memory and/or cognitive control at which the intended structure for the unambiguous sentence is always found, but the intended structure for the temporarily-ambiguous sentence is not. By parameterizing working memory and cognitive control, we formalize the role of these two extralinguistic systems in sentence parsing across development.

Working memory is always limited in the parser, even for adults (Caplan & Waters, 1999; Just & Carpenter, 1992; Lewis, 1996; O'Rourke, 2013), which produces the risk that any given partial analysis can get 'lost'. This 'loss' is thought to be the source of revision difficulties in garden path sentences (Jurafsky, 1996). In children, this type of loss may be accentuated by two systems that are still developing. When working memory limitations are more severe, partial analyses are subject to stricter score thresholds, meaning that if a particular candidate has a relatively lower score at any point during parsing, that candidate may be lost. Alternatively, a lag in score-updating via immature cognitive control could cause some candidates to maintain artificially high scores, swamping the memory allocated to pursuing the other analyses and causing them to fall off the beam. In Simulation 1, we found that both immature cognitive control and developing working memory could be individually responsible for children's parsing performance in Trueswell et al.

(1999). We found that at many combinations of *k* and *w* where the intended analysis is not found, the parser finds a structure analogous to the "kindergarten path" analysis (Figure 6.4), with two disconnected fragments (a *VP* and a *PP*) rather than one connected structure for the entire string.

In Simulation 2, we found that the two possible sources of children's parsing difficulties – working memory and cognitive control – make different predictions when it comes to handling a case of global ambiguity in the Trueswell et al. (1999) test sentence that has not been considered experimentally. We found that at values of *k* and *w* where the parser cannot find the intended analysis for the test scene, it finds the alternative analysis of this sentence where *the napkin in the box* is the destination for the frog. This suggests that children's seeming inability to interpret the test sentence in Trueswell et al. (1999) could arise not as a result of failing to find one connected structure, but rather due to competition between this alternative analysis and the intended analysis.

By varying the parameters corresponding to the influence of working memory and cognitive control incrementally, we generated testable empirical predictions which could help determine whether limited working memory alone or developing cognitive control could be responsible for the differences between children's and adults' parsing. Namely, the results from Simulation 2 suggest that if a developing cognitive control system impacts parsing through delayed score-updating, we would expect children to go through two developmental stages. At earlier stages, when immature cognitive control results in slower score updating and artificially-high scoring candidates are pursued, only the alternative analysis will be found. At later stages of development, when cognitive control interacts with sentence processing in an adult-like way and allows for faster score updating, both the intended and alternative analyses will be accessible. If, however, working memory is solely responsible for the kindergarten path effect, our parser makes a different prediction with three developmental stages. Namely, at earlier stages of development when working memory is more restricted, our parser predicts that only the alternative analysis will be found. However, as working memory develops and interacts with sentence processing in a more adult-like way, this stage will be followed by another stage in which *only* the originally intended analysis can be accessed. Lastly, as working memory approaches an adult-like state and becomes even less restrictive, both analyses will become accessible. Thus, our model generates testable predictions to adjudicate between these two possibilities.

These predictions could be tested with a new experimental design that reveals whether children can find the alternative parse of the sentence. This could be achieved with an additional condition, with a layout similar to that of the original study. Rather than having two candidate frogs to be moved (with one on a napkin), in this layout there would be a frog, a distractor animal, one napkin in its own quadrant, and a napkin within a box (see Figure 6.12 for an example setup). Children (and adults) would be prompted with the same ambiguous test sentence: *Put the frog on the napkin in the box.* The unambiguous test sentence, however, would be *Put the frog on the napkin that is in the box*.

Figure 6.12: Example set-up for testing the *on*-as-goal analysis.

To test the predictions of our model, this design would need to include both a condition with this novel layout, and another with the original layout. Our model predicts that if immature working memory, and not immature cognitive control, is responsible for children's parsing difficulties in Trueswell et al. (1999), there will be a stage of development where children can successfully find the parse of the ambiguous sentence that is supported by this layout (*on*-as-goal), but not the intended parse in the original Trueswell et al. (1999) layout (*on*-as-modifier). This would be demonstrated by children successfully acting out the sentence in the novel layout condition (Figure 6.12), placing the frog in the upper left quadrant onto the napkin in the box in the upper right quadrant, while failing to correctly act out the sentence as intended in the original layout condition. This developmental stage would be followed by another, where children would only be able to find the originally intended parse for the Trueswell et al. (1999) scene (*on*-as-modifier) and not for this alternative scene. They would consequently be unable to act out the ambiguous sentence in this alternative layout, as there is no frog on the napkin. Crucially, the same children would succeed in the original layout, successfully moving the frog that is on the napkin into the box. Lastly, as children approach maturity, they would be able to act out all sentences as intended when presented with either scene.

If, on the other hand, immature cognitive control is responsible for children's parsing difficulties, we would not expect any such trade-off between the two analyses: children would only be able to find the alternative analysis at earlier points in development, followed by a period where children could successfully find both analyses. At earlier stages, children would only be able to act out the sentence in the novel layout condition, but not the original layout condition. At later stages, children would be able to successfully act out the sentence as intended in both conditions. This novel adaptation of Trueswell et al. (1999)'s study could thus be used to adjudicate between the two developing cognitive systems as the source of children's difficulties.

The question arises of how we can interpret cases where multiple analyses are found by our parser, but the wrong (i.e., alternative) analysis is found first. It is plausible that these cases are akin to mild garden-path effects found in the adult literature in which the intended analysis is found but after a slight delay, that occurs when the parser arrives at an analysis that seems wrong or dispreferred in the context (either because it isn't one connected structure, or it is incompatible with the visual context). Once this incongruity is encountered, the parser is prompted to backtrack to another analysis that is still available. If this is the case, what remains on the beam corresponds to candidates that the parser can backtrack to, without having to start over and re-analyze the sentence (Hale, 2011). If this type of backtracking is occurring, we would expect children in these developmental stages to be able to recover from the initial mild garden-pathing and find the intended analysis. Alternatively, if children are unable to backtrack, we would expect children to be unable to find the intended analysis, behaving as they would in a true "kindergarten pathing". We leave adjudication between these two possibilities to future investigation.

6.6.1 Extensions of the model

One factor that we know contributes to children's parsing behavior with *Put the frog on the napkin in the box* in Trueswell et al. (1999)'s study is that children's knowledge of verb argument structure leads them to commit to a parse for the sentence that turns out to be incompatible with the context. Namely, children's knowledge that *put* requires both an *NP* argument and a *PP* argument leads them to make an initial commitment to an analysis in which *the frog* constitutes the *NP* and *on the napkin* constitutes the goal *PP*. Although the current model only focused on one case of parsing difficulty due to argument structure knowledge, it could be extended to other cases examining how developing argument structure interacts with the other developing cognitive systems we have formally parameterized.

For example, the parser could be extended to model the findings of Lidz et al. (2017), in which 16 month-olds successfully identify *the tig* as the instrument in the sentence *She's wiping with the tig*, while 19 month-olds incorrectly learn that *the tig* is the patient. As discussed in Chapter 5, this discrepancy likely arises as a result of 19 month-olds' knowledge of the verb *wipe*: because this verb occurs more frequently with a direct object than a prepositional object, 19 month-olds make the prediction that *the tig* is the direct object. This could be viewed as a form of garden-pathing: an initial prediction leads to an incorrect analysis of the sentence. On the other hand, 16 month-olds successfully learn that *the tig* is the instrument because they do not exploit knowledge of *wipe*'s argument structure to make the prediction that the following *NP* is the patient, relying on bottom-up information in the sentence instead.

To model the findings of Lidz et al. (2017), the PCFG fed to the model could be modified to reflect different stages of verb argument structure acquisition. Because this grammar is meant to model the linguistic knowledge of a child at a particular point in development, the grammar G could be modified to reflect the child's estimated vocabulary knowledge by referencing typical vocabulary inventories (i.e., mCDI scores, Dale and Fenson, 1996) for children of different ages. Any lexical items that have not yet been acquired could be collapsed into an "unknown verb" category. This category would reflect the overall weights of expansions of VP, including argument structure, across all of the verbs that children at these ages do know. These expansions could be weighted equally to reflect all of the possible argument structures that the child has acquired. Thus, if the verb *wipe* has not yet been acquired, the parser using an appropriately-modified PCFG might consider any possible VP argument structure in its grammar as a viable option. By collapsing these categories into the unknown group, we anticipate that our model, at appropriate (i.e., more restrictive) settings of k and/or w, will be able to capture the finding that 16 month-olds without prior knowledge of the argument structure of *wipe* will succeed in parsing an example like *She's wiping with the tig*, while 19 month-olds with the relevant knowledge would be garden-pathed. Recent work from White and Lidz (2022) found that 19 month-olds tested on the same paradigm, but with a novel verb, once again succeeded at discerning the patient and instrument labels. In the case of novel verbs, 19 month-olds do not generate predictions on the basis of verb argument structure, leading to no issues of revision. These findings suggest that children's parsing predictions are based on their lexically-specific verb argument structure knowledge, which we could capture through modifying the PCFG. Through similar manipulations of the PCFG, the parser could also be used to generate novel predictions for children's parsing performance with previously untested verbs.

Our parser can further be extended to model cognitive control as a resource that can be up- or down-regulated, rather than a trait. For example, previous work has shown that when cognitive control is engaged for adults (by introducing incongruity in preceding trials, for example), their performance on garden-path sentences improves (Novick et al., 2005). On the other hand, children tasked with conflict resolution immediately preceding a garden-path sentence trial performed worse (Ovans, 2022). This discrepancy suggests that the cognitive control system may be a resource, recruited in different ways across development. Extending our model to this sort of resource-based model of cognitive control, w could vary within a range, and another parameter could control what value it takes on within that range. Development could then be modeled as an expansion of this range. Our model lays the groundwork for future, more nuanced explorations.

In summary, the parser implemented in this chapter sets the stage for a wide variety of further computational and experimental work, allowing the research community to formalize and pinpoint the exact mechanism(s) by which working memory and immature cognitive control can influence children's deployment of verb argument structure knowledge and any other acquired linguistic knowledge that forms the basis of incremental sentence processing.

CHAPTER 7

Conclusion

In this dissertation, we investigated the role of the extralinguistic cognitive systems that are developing in tandem with the linguistic system in the acquisition and later deployment of verb argument structure knowledge. Understanding the precise ways in which these systems influence learning is crucial for developing robust theories of language acquisition and development, as well as pinpointing how children learn to use the language they are so rapidly acquiring.

7.1 Summary of key findings

7.1.1 Extralinguistic cognition in argument structure learning

In Part 1 of the dissertation, we focus on the learning of new verbs, including their meaning and argument structure, and investigate the perceptual support afforded by developing extralinguistic perceptual systems. Understanding how children represent the events they see in the world around them is crucial for the development of theories of how children learn verbs. Given that the events unfolding in the world are multi-dimensional, and can be described linguistically in (virtually) unlimited ways, verb learning is no small task. Bootstrapping theories offer one solution, positing that children can exploit robust correlations between meaning and syntax to link their conceptual representation of a scene to the linguistic label being heard. But understanding the exact mechanisms that drive this linking during early verb learning requires us to pinpoint the non-linguistic conceptual representations under which children view the scenes around them.

We approached this question with the case study of *trade*, one of only a handful of simple verbs across the world's languages that is licensed to occur with 4 arguments. This predicate typically labels an event concept that plausibly has four participants: two individuals mutually exchanging two items. Representing all four of these participants could pose a challenge to the young learner, however: at 14 months, infants are reported to have a visual working memory limit of only three items (Feigenson & Carey, 2003; Feigenson & Halberda, 2008; Stahl et al., 2023). Consequently, probing the nature of the conceptual representations that could support the young learner's acquisition of verbs like *trade* has fundamental implications for our theories of verb-learning. Any theory of language acquisition must provide plausible strategies for the learner to link meaning and structure even in the face of developing cognitive systems outside of language.

We began by probing adults' conceptual representations of a trading scene in Chapter 3. We found that adults explicitly encode all four participants – both traders and both items traded – in their conceptual representations of a trading scene. Moreover, by comparing their representations of another plausibly four-participant scene, a giving-then-disposing, we provided converging evidence that adults view this trading scene under a single event concept, rather than two. Our study provides some of the first evidence for the non-linguistic conceptual representations of such high-adicity events, and suggests that even though the adult visual working memory system typically has a limit of 4 items (Cowan, 2001; Halberda et al., 2006; Scholl & Pylyshyn, 1999; Sperling, 1960; Trick & Pylyshyn, 1993), representing these four participants does not exceed that capacity. We then turned to the question of whether preschool-aged children can similarly represent all four participants in a trading scene in Chapter 4. We found that children are able to represent both traders and both items-traded in the same trading scene, suggesting that despite more stringent visual working memory limitations, children aged 3.5-5.5 years are still able to represent this scene under a 4-participant concept. A brief survey of English child-directed-speech indicated that *trade* rarely occurs in 4-argument frames (e.g., *Dan and Sue traded a truck and a ball*), occurring more frequently in frames with 1 or 2 collective arguments instead (e.g., *They traded (their) toys*). This finding raises questions about the specific strategies that young children can deploy to learn verbs like *trade*.

Our findings have implications for the theories regarding how children link between their conceptual and linguistic representations. Namely, by investigating the non-linguistic conceptual representations of complex event types, we highlight the importance of considering children's extralinguistic support for language learning. Our findings that children represent all four participants in a trading scene set the groundwork for further work testing different theories of how children link between their conceptual and linguistic representations for these event types. By further probing the internal structure of children's representations of events, future work could help adjudicate between two classes of linking theories for arguments and participants. If, for example, children rely on one-to-one matching of linguistic arguments to participants (Fisher, 1996; Lidz & Gleitman, 2004; Naigles, 1990; Yuan et al., 2012), the overall rarity of *trade* occurring in a 4-argument frame could present substantial difficulties if, like adults, children also represent tradings under a 4-participant concept. This linking strategy may not pose as much of a challenge if the event is instead represented as two 3-participant GIVINGs, given that *trade* frequently occurs in 3-argument frames and could consequently lend itself to this type of one-to-one mapping. If, on the other hand, children rely on links between thematic roles and participants (Baker & Levin, 2015; Dowty, 1991; He, 2015; Jackendoff, 1990; Perkins et al., 2024; Williams, 2015), viewing the scene under either a single 4-participant concept or two 3-participant concepts, should not pose substantial difficulties even when infrequently hearing *trade* in a 4-argument frame, provided children can link the syntax they are hearing to the scene representation in other ways.

This work further presents the opportunity to expand our current theories of bootstrapping to accommodate the acquisition of high-adicity verbs. For example, the study of high-adicity verbs like *trade* could point to additional linking strategies that children could exploit to learn complex verbs: perhaps, children could exploit the knowledge that predicates that exist in collective-reciprocal alternations tend to identify symmetrical events in order to acquire high-adicity verb meanings.

7.1.2 Extralinguistic cognition in argument structure processing

In Part 2 of the dissertation, we focused on the deployment of verb argument structure knowledge, focusing on the ways that developing cognitive systems can influence how children apply existing knowledge about verbs during online sentence processing. Children, and even infants, deploy their knowledge of verb argument structure to make early commitments and predictions about upcoming linguistic information (Lidz, 2023; Lidz et al., 2017; Omaki & Lidz, 2015; Trueswell et al., 1999, among others). However, although children's parsing is incremental and predictive just like that of adults, their parsing differs from that of adults' in the seeming inability to revise from initial commitments. It has been hypothesized that developing extralinguistic cognitive systems, like working memory and cognitive control, are responsible for the seeming inability of children to recover from erroneous initial commitments. We implemented a computational parser that models the development of each of these two systems individually and in tandem, allowing us to pinpoint the precise ways that these systems can intersect with processing. This type of modeling is crucial to develop a unified theory of sentence processing across development, accommodating not only how language knowledge is deployed at adult stages of cognitive development, but also at the very earliest stages of language learning. Moreover, understanding how immature cognitive systems can influence early sentence understanding, potentially masking underlying linguistic competency, is further crucial for the interpretation of experimental and behavioral results in the field of language acquisition research.

In Chapter 6, we implemented an incremental and predictive parser that models the influence of cognitive control and working memory in sentence processing across development. We parameterized each system according to prior hypotheses on how they interact with parsing. Namely, to model the influence of working memory on parsing, our working memory parameter constrained the number of alternatives maintained at each parsing choice point, keeping only the most probable analyses at immature settings of the parameter and allowing additional analyses to persevere at more mature settings. Our cognitive control parameter effectively delayed the score updating of a given analysis, such that at immature settings of the parameter, if an analysis had previously been strongly favored, it could persist for some time even in the face of conflicting evidence. At mature settings of the cognitive control parameter, however, analyses were updated immediately if conflicting evidence was encountered.

By formalizing the role of each system as independently adjustable parameters, we were able to investigate the role of each developing system as it approaches maturity. We conducted two simulations exploring the source of children's difficulty with the temporarily-ambiguous sentence in Trueswell et al. (1999), *Put the frog on the napkin in the box*. In Simulation 1, we demonstrated that children may entertain a disconnected analysis, in which the PP *in the box* is not incorporated into the structure and the sentence is interpreted as *Put the frog on the napkin* based on their initial prediction that *the napkin* is the destination for the frog. This analysis outcompetes the intended analysis, (where *on the napkin* modifies *the frog* and *in the box* is the destination for the movement) when either of the parameters corresponding to working memory and cognitive control are set to child-like values. In Simulation 2, we examined the competition between the intended analysis and a previously untested alternative analysis resulting from the global ambiguity of the test sentence. We found that the parser makes different predictions for this alternative analysis, in which *the napkin in the box* is the destination for the frog, for the two previous hypotheses about the roles of each developing system. Our parser predicts that children will initially only be able to find the alternative analysis. If the immaturity of cognitive control is the culprit for children's processing difficulties, our parser predicts that as this system develops, children will enter a phase of development where they can find both analyses in an adult-like way. On the other hand, if cognitive control does not interact with processing in the hypothesized manner and working memory is the sole culprit, our parser makes a different prediction: children will initially only be able to find the alternative analysis, followed by a period in which they can only find the originally intended analysis. Finally, as the working memory system approaches maturity, children will be able to find both analyses. We proposed a novel experimental paradigm that could test this developmental prediction to tease apart the role of each system within the developing parser.

7.2 Future directions

This dissertation laid the foundational groundwork for future research. Our findings raise several questions that we leave to future work to address.

Our findings in Part 1 centered around the non-linguistic representations of complex event types, such as *trade*. This work invites future research examining the detailed internal structure of the event representations under which the test scene was viewed. For example, we found that adults represent all four participants in the trading scene under a single concept. Although we ruled out the possibility of adults viewing this scene as two sequential GIVINGs, this finding leaves open the possibility of further internal structure: for example, participants could be 'chunked' by participant type (traders and items traded) or by initial possession. Similarly, we found that children are capable of representing all four participants in a trading scene. However, it is still unknown whether children view the trading scene under a single event representation with four participant relations, similar to what we found in adults, or whether they represent the event as two sequential GIVINGs due to limitations imposed by their developing working memory systems. Additional future work could address this possibility, as well as investigate whether the representations of children contain internal chunking structure.

Our work begins to answer an important question in early verb learning by identifying the non-linguistic conceptual representation of high-adicity event type. Future work could more directly probe the types of verb learning strategies that could be applied to map between the representation of such a complex event and the syntactic structure of the sentence in which the verb occurs. Additionally, further cross-linguistic analysis of verbs like *trade* could potentially elucidate aspects of the linguistic representation that could map to non-linguistic conceptual representations. For example, if *trade* is inherently reciprocal in other languages, there may be morphological evidence of this reciprocality which could be exploited by learners of those languages. We leave this type of inquiry to future work.

In Part 2, we developed one of the first formally explicit models of how children's developing linguistic knowledge interacts with two developing extralinguistic cognitive abilities in sentence processing. While our parser focused on the results of Trueswell et al. (1999) as a case study, this work lays the foundational groundwork for additional experimental research and for computational work. Our model's empirical predictions provide an immediate avenue for experimental investigations of the two systems that may be responsible for the difference between adults' and children's sentence processing. Beyond this, our model provides a foundation which could be extended to investigate other questions about children's parsing in other contexts.

As a starting point, we chose to model cognitive control as a trait characteristic, rather than as a resource that can be recruited. However, it is likely that cognitive control can be engaged to greater or lesser extents within the same individual (Hagger et al., 2010; Ness et al., 2022; Ovans, 2022; Powell & Carey, 2017). Future work could attempt to model a more dynamic system by modifying the cognitive control parameter to vary throughout parsing. Adapting this model in such a way may also capture the differences between children's and adult's performance with ambiguous sentences when cognitive control is engaged (Ovans, 2022). As such, we leave our parser as the foundational groundwork for more sophisticated investigation of the role of cognitive control in sentence processing through development.

Our parser models the influence of extralinguistic cognition on sentence processing, focusing purely on predictive and incremental parsing based in linguistic knowledge. As discussed in Chapter 5, sentence processing in adults can be guided not just by linguistic knowledge, but also world knowledge, visual context, and linguistic context (Altmann & Steedman, 1988; Tanenhaus et al., 1995; Trueswell et al., 1999, among others). In future work, our parser could be expanded to accommodate sensitivity to these other forms of knowledge. To accommodate a developmental trajectory, this model would necessarily need to reflect children's seeming *in*sensitivity to these extra-linguistic forms of knowledge, gradually allowing this type of information to become more informative as the parser matures.

7.3 Conclusion

This dissertation investigated a crucial underlying property of language learning and development: the learning process interacts with still-developing cognitive systems outside of the linguistic system. By investigating these independently developing cognitive systems through the lens of verb argument structure, we explored two ways that these systems can influence verb learning and use. In doing so, we provide an example of how to isolate the specific contributions of extralinguistic cognition to the acquisition and deployment of one type of linguistic knowledge.

This type of approach, which combines computational modeling with psychological evidence for how children represent the extralinguistic contexts that are used for acquiring linguistic meaning, allows for the development of formally precise hypotheses regarding how children deploy linguistic knowledge in real-time with the support of their extralinguistic cognitive systems. Such formally precise hypotheses are crucial for the development of robust theories of language development, as language learning is an iterative process which continually builds upon existing (and imperfect) linguistic and extralinguistic knowledge (Fodor, 1998; Lidz & Gagliardi, 2015; Perkins et al., 2022; Valian, 1990). Because the extralinguistic cognitive systems supporting early language learning are still developing in and of themselves over the course of early language learning, this type of investigation is crucial to pinpoint the types of representations children learn from as they approach the fundamental challenge of language acquisition. For example, if an immature visual working memory system provides inadequate support for conceptual representations of complex event types, this could pose additional challenges for mapping to their linguistic representations of verbs that label those event types. Similarly, if an immature cognitive control system prevents a child from parsing their input veridically, they may be led to incorrect generalizations as they attempt to learn from incomplete linguistic representations of the sentences spoken around them (e.g., Fodor, 1998; Lidz et al., 2017). Non-linguistic evidence of the type discussed in this dissertation is crucial for refining our hypotheses about language learning and use in development and extends to aspects of language development beyond the case studies that are considered here.

APPENDIX A

Picky puppet script

Experimenter: Today we're going to meet my friend named Miss Hippo, and she's a very picky hippo. She only likes it when things are the same, and she *really* doesn't like it when things are different. Are you ready to meet Miss Hippo?

Experimenter pulls out the puppet.

Miss Hippo: Hi! I'm Miss Hippo. I'm a very picky hippo! I *only* like it when the same thing happens, and I really don't like it when something different happens. What are we doing today?

Experimenter: Well, today we're going to watch a bunch of videos together. I've got a list of all of the videos we're going to watch right here. Each time, we're going to see two videos. Our job is to watch the first video and the second video and see if the same thing happens. If the same thing happens, we're going to put a happy sticker down, because that makes Miss Hippo happy! And if something different happens, we're going to put a sad sticker down, because that makes Miss Hippo sad. Are you ready to watch the first video together?

Experimenter directs attention to the laptop and plays the first training trial.

Miss Hippo: Yay!! I like that the same thing happened in those two videos! Happy sticker for that one, the same thing happened!

Experimenter prompts child to place a happy sticker down, then plays the next training trial.

Miss Hippo: Yay!!! I like that the same thing happened in those two videos, too! Another happy sticker for that one, the same thing happened!

Experimenter prompts child to place a happy sticker down, then plays the next training trial.

Miss Hippo: Yucky!!! I don't like that something different happened in those two videos, yucky yuck! Sad sticker for that one, I *only* like it when the same thing happens!

Experimenter prompts child to place a sad sticker down, then plays the next training trial.

Miss Hippo: Yucky yuck!!! I don't like that something different happened in those two videos, yuck! Sad sticker for that one, too, I *only* like it when the same thing happens!

Experimenter prompts child to place a sad sticker down. Because the child has now watched four videos, a spinning star appears on the screen; experimenter prompts child to put the "special sparkly star" sticker on the additional sheet of paper.

Miss Hippo: *(Yawns loudly.)* I'm getting pretty tired – I haven't had my nap yet today. I'm so tired I'm might fall asleep right here! *(Yawns again, then "falls asleep" on the table.)*

Experimenter: Can you believe it? Miss Hippo fell asleep in the middle of the activity! Do you think we can watch the next two videos and see if the same thing happens before she wakes up?

Experimenter plays the first practice trial, then prompts child for a response. After the response is given, experimenter "wakes up" Miss Hippo.

Miss Hippo: Oh, I'm so sorry, I fell asleep! Can we watch those videos one more time, so I can see what happened?

Experimenter plays the first practice trial again, allowing Miss Hippo to 'watch'.

Miss Hippo: Yay!! I like that the same thing happened in those two videos! You're right, happy sticker for this one, happy sticker because the same thing happened! *(Yawns again)* I really am pretty tired. I might even fall asleep again *(yawn)* right here...! *(Falls asleep on table again.)*

Experimenter: Can you believe it? Miss Hippo fell asleep *again*! She's very sleepy. Do you think we can watch the next two videos and see if the same thing happens before she wakes up again?

Experimenter plays the second practice trial, then prompts child for a response. After the response is given, experimenter "wakes up" Miss Hippo.

Miss Hippo: Yucky!! I don't like that something different happened in those two videos! You're right, sad sticker for this one, sad sticker because something different happened! I *only* like when they're the same. *(Yawns again)* I really am pretty tired. Do you think I could go take a nap? Can you finish watching the videos and putting down the stickers for me?

Bibliography

Altmann, G. (1986). Reference and the resolution of local syntactic ambiguity: The effect of context during human sentence processing.

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

Anderson, P. (2002). Assessment and development of executive function (ef) during childhood. *Child neuropsychology*, *8*(2), 71–82.

Arunachalam, S., & Waxman, S. R. (2010). Meaning from syntax: Evidence from 2-year-olds. *Cognition*, *114*(3), 442–446.

Baddeley, A. (1992). Working memory [Publisher: American Association for the Advancement of Science]. *Science*, *255*(5044), 556–559. Retrieved March 12, 2024, from https://www.jstor.org/stable/2876819

Baddeley, A. (2003). Working memory: Looking back and looking forward [Number: 10 Publisher: Nature Publishing Group]. *Nature Reviews Neuroscience*, *4*(10), 829–839. https://doi.org/10.1038/nrn1201

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.). Academic Press. https://doi.org/https://doi.org/10.1016/S0079-7421(08)60452-1

Baker, L. J., & Levin, D. T. (2015). The role of relational triggers in event perception. *Cognition*, *136*, 14–29.

Bayliss, D. M., Jarrold, C., Baddeley, A. D., Gunn, D. M., & Leigh, E. (2005). Mapping the developmental constraints on working memory span performance. *Developmental psychology*, *41*(4), 579.

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, *132*(1), 71.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In M. Sanz, I. Laka, & M. K. Tanenhaus (Eds.), *Language down the garden path* (pp. 1–80). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199677139.003.0001

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.

Botvinick, M. M., Carter, C. S., Braver, T. S., Barch, D. M., & Cohen, J. D. (2001). Conflict monitoring and cognitive control.

Bowerman, M. (1974). Learning the structure of causative verbs: A study in the relationship of cognitive, semantic, and syntactic development. *Papers and reports on child language development*, *8*, 142–178.

Bowerman, M. (1977). The acquisition of rules governing 'possible lexical items': Evidence from spontaneous speech errors. *Papers and Reports on Child Language Development*, *13*, 148–156.

Bowerman, M. (1982). Evaluating competing linguistic models with language acquisition data: Implications of developmental errors with causative verbs. *Quaderni di semantica*, *3*, 5–66.

Brants, T., & Crocker, M. (2000). Probabilistic parsing and psychological plausibility. *COLING 2000 Volume 1: The 18th International Conference on Computational Linguistics*. Retrieved June 23, 2023, from https://aclanthology.org/C00-1017

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in cognitive sciences*, *16*(2), 106–113.

Brosnan, S. F., & De Waal, F. B. (2002). A proximate perspective on reciprocal altruism. *Human Nature*, *13*, 129–152.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*(1), 77–94. https://doi.org/10.1017/S0140525X99001788

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of experimental child psychology*, *33*(3), 386–404.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive psychology*, *4*(1), 55–81.

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT press.

Chuah, Y. L., & Maybery, M. T. (1999). Verbal and spatial short-term memory: Common sources of developmental change? *Journal of Experimental Child Psychology*, *73*(1), 7–44.

Cohen, J. D., & Huston, T. A. (1994). Progress in the use of interactive models for understanding attention and performance.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, *24*(1), 87–114.

Crocker, M. W., & Brants, T. (2000). Wide-coverage probabilistic sentence processing. *Journal of Psycholinguistic Research*, *29*(6), 647–669. https://doi.org/10.1023/A:1026560822390

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, *19*(4), 450–466.

Daneman, M., & Tardif, T. (2016). Working memory and reading skill re-examined. In *Attention and performance xii* (pp. 491–508). Routledge.

De Waal, F. B. (1989). Food sharing and reciprocal obligations among chimpanzees. *Journal of Human Evolution*, *18*(5), 433–459.

Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental psychology*, *38*(3), 352.

Doebel, S., & Zelazo, P. D. (2015). A meta-analysis of the dimensional change card sort: Implications for developmental theories and the measurement of executive function in children. *Developmental Review*, *38*, 241–268. https://doi.org/10.1016/j.dr.2015.09.001

Dowty, D. (1991). Thematic proto-roles and argument selection. *language*, *67*(3), 547–619.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of memory and language*, *27*(4), 429–446.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 972.

Enloe, J. (2003). Food sharing past and present: Archaeological evidence for economic and social interactions. *Before Farming*, *2003*(1), 1–23.

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science*, *6*(5), 568–584.

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition*, *97*(3), 295–313.

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological science*, *13*(2), 150–156.

Feigenson, L., & Halberda, J. (2004). Infants chunk object arrays into sets of individuals. *Cognition*, *91*(2), 173–190.

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences*, *105*(29), 9926–9930.

Feistner, A. T., & McGrew, W. C. (1989). Food-sharing in primates: A critical review. *Perspectives in primate biology*, *3*(21-36).

Fillmore, C. J. (1968). Lexical entries for verbs. *Foundations of language*, 373–393.

Fillmore, C. J. (1972). Subjects, speakers, and roles. In *Semantics of natural language* (pp. 1–24). Springer.

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). Labvanced: A unified javascript framework for online studies. *International conference on computational social science (cologne)*, 1–3.

Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, *31*(1), 41–81. https://doi.org/10.1006/cogp.1996.0012

Fodor, J. D. (1998). Parsing to learn. *Journal of Psycholinguistic research*, *27*, 339–374.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies.* (Doctoral dissertation) [ISBN: 9798661175977]. University of Connecticut. United States – Connecticut. Retrieved March 12, 2024, from https://www.proquest.com/docview/302925499/citation/CF1A35DA23494818PQ/1

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 312–7 years old on a stroop-like day-night test. *Cognition*, *53*(2), 129–153.

Gibson, E., et al. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, *2000*, 95–126.

Gleitman, L. (1990). The structural sources of verb meanings. *Language acquisition*, *1*(1), 3–55.

Gleitman, L. R., Gleitman, H., Miller, C., & Ostrin, R. (1996). Similar, and similar concepts.

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in cognitive sciences*, *5*(6), 236–243.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of child language*, *14*(1), 23–45.

Gordon, P. (2003). The origin of argument structure in infant event representations. *Proceedings of the 28th Annual Boston University Conference on Language Development*, *1*, 189–198.

Gordon, P., & Chafetz, J. (1990). Verb-based versus class-based accounts of actionality effects in children's comprehension of passives. *Cognition*, *36*(3), 227–254.

Grimshaw, J. (1981). Form, function, and the language acquisition device. *The logical problem of language acquisition*, *165*, 178.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The learnability and acquisition of the dative alternation in english. *Language*, 203–257.

Gurven, M. (2004). To give and to give not: The behavioral ecology of human food transfers. *Behavioral and Brain Sciences*, *27*(4), 543–560.

Hafri, A., & Firestone, C. (2021). The perception of relations. *Trends in Cognitive Sciences*, *25*(6), 475–492. https://doi.org/10.1016/j.tics.2021.01.006

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, *142*(3), 880.

Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition*, *175*, 36–52.

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological bulletin*, *136*(4), 495.

Halberda, J., Simons, D., & Wetherhold, J. (2006). You can never attend to more than three items at once: Gestalt grouping principles explain changes in capacity. *Manuscript submitted for publication*.

Hale, J. T. (2011). What a rational parser would do [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1551 6709.2010.01145.x]. *Cognitive Science*, *35*(3), 399–443. https://doi.org/10.1111/j.1551-6709.2010.01145.x

He, X. (2015). *Verb learning under guidance* (Doctoral dissertation). University of Maryland, College Park.

Henderson, J. (2004). Lookahead in deterministic left-corner parsing. *Proceedings of the workshop on incremental parsing: Bringing engineering and cognition together*, 26–33.

Henriksen, I., Bilardi, G., & Pingali, K. (2019). Derivative grammars: A symbolic approach to parsing with derivatives. *Proceedings of the ACM on Programming Languages*, *3*, 127:1–127:28. https://doi.org/10.1145/3360553

Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *language*, 251–299.

Hsu, N. S., Kuchinsky, S. E., & Novick, J. M. (2021). Direct impact of cognitive control on sentence processing and comprehension. *Language, Cognition and Neuroscience*, *36*(2), 211–239.

Hsu, N. S., & Novick, J. M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing [Publisher: SAGE Publications Inc]. *Psychological Science*, *27*(4), 572–582. https://doi.org/10.1177/0956797615625223

Jackendoff, R. (1990). What would a theory of language evolution have to look like? *Behavioral and Brain Sciences*, *13*(4), 737–738.

Jackendoff, R. S. (1992). *Semantic structures* (Vol. 18). MIT press.

Jaeggi, A. V., Burkart, J. M., & Van Schaik, C. P. (2010). On the psychology of cooperation in humans and other primates: Combining the natural history and experimental evidence of prosociality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1553), 2723–2735.

January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of stroop and syntactic ambiguity resolution in broca's area: Implications for the neural basis of sentence processing. *Journal of cognitive neuroscience*, *21*(12), 2434–2444.

Jin, K.-s., & Fisher, C. (2014). Early evidence for syntactic bootstrapping: 15-month-olds use sentence structure in verb learning. *BUCLD 38 Online Proceedings Supplement*.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog2002_1]. *Cognitive Science*, *20*(2), 137–194. https://doi.org/10.1207/s15516709cog2002_1

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*(1), 15–47. https://doi.org/10.1016/0010-0277(72)90028-5

Landau, B., & Gleitman, L. (1985). *Language and experience: Evidence from the blind child* (Vol. 8). Harvard University Press.

Leslie, A. M. (1982). The perception of causality in infants. *Perception*, *11*(2), 173–186.

Leslie, A. M. (1984). Spatiotemporal continuity and the perception of causality in infants. *Perception*, *13*(3), 287–305.

Leslie, A. M., Xu, F., Tremoulet, P. D., & Scholl, B. J. (1998). Indexing and the object concept: Developingwhat'andwhere'systems. *Trends in cognitive sciences*, *2*(1), 10–18.

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of psycholinguistic research*, *25*(1), 93–115.

Lidz, J. (2023). Parser-grammar transparency and the development of syntactic dependencies. *Language Acquisition*, *30*(3-4), 311–322.

Lidz, J., & Gagliardi, A. (2015). How nature meets nurture: Universal grammar and statistical learning. *Annu. Rev. Linguist.*, *1*(1), 333–353.

Lidz, J., & Gleitman, L. R. (2004). Argument structure and the child's contribution to language learning. *Trends in Cognitive Sciences*, *8*(4), 157–161. https://doi.org/10.1016/j.tics.2004.02.005

Lidz, J., White, A. S., & Baier, R. (2017). The role of incremental parsing in syntactically conditioned word learning. *Cognitive Psychology*, *97*, 62–78.

Lowerre, B. P., & Reddy, B. R. (1976). Harpy, a connected speech recognition system. *The Journal of the Acoustical Society of America*, *59*(S1), S97–S97.

MacDonald, M. C. (1997). Lexical representations and sentence processing: An introduction [Publisher: Routledge _eprint: https://doi.org/10.1080/016909697386826]. *Language and Cognitive Processes*, *12*(2), 121–136. https://doi.org/10.1080/016909697386826

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological review*, *101*(4), 676.

MacWhinney, B. (2000). *The childes project: The database* (Vol. 2). Psychology Press.

Maratsos, M., Gudeman, R., Gerard-Ngo, P., & DeHart, G. (2014). A study in novel word learning: The productivity of the causative. In *Mechanisms of language acquisition* (pp. 89–113). Psychology Press.

Meeren, H. K., Van Heijnsbergen, C. C., & De Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences*, *102*(45), 16518–16523.

Milham, M., Banich, M., Webb, A., Barad, V., Cohen, N., Wszalek, T., & Kramer, A. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research*, *12*(3), 467–473.

Miller, E. K. (2000). The prefontral cortex and cognitive control. *Nature reviews neuroscience*, *1*(1), 59–65.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, *24*(1), 167–202.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

Moher, M., Tuerk, A. S., & Feigenson, L. (2012). Seven-month-old infants chunk items in memory. *Journal of experimental child psychology*, *112*(4), 361–377.

Montgomery, J. W., Magimairaj, B. M., & O'Malley, M. H. (2008). Role of working memory in typically developing children's complex sentence comprehension. *Journal of Psycholinguistic Research*, *37*(5), 331–354. https://doi.org/10.1007/s10936-008-9077-z

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of child language*, *17*(2), 357–374.

Nederhof, M.-J. (1993). Generalized left-corner parsing. *Sixth Conference of the European Chapter of the Association for Computational Linguistics*. Retrieved June 7, 2023, from https://aclanthology.org/E93-1036

Ness, T., Langlois, V. J., Kim, A. E., & Novick, J. (2022). The state of cognitive control in language processing.

Noble, C. H., Rowland, C. F., & Pine, J. M. (2011). Comprehension of argument structure and semantic roles: Evidence from english-learning children and the forced-choice pointing paradigm. *Cognitive science*, *35*(5), 963–982.

Nordlinger, R. (2023). The Typology of Reciprocal Constructions. *Annual Review of Linguistics*, *9*(1), 71–91. https://doi.org/10.1146/annurev-linguistics-022421-064006

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In *Consciousness and self-regulation: Advances in research and theory volume 4* (pp. 1–18). Springer.

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience*, *29*(2), 186–217.

Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex [Publisher: Routledge _eprint: https://doi.org/10.1080/02643290903519367]. *Cognitive Neuropsychology*, *26*(6), 527–567. https://doi.org/10.1080/02643290903519367

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(3), 263–281.

Nowak, F., & Cotterell, R. (2023). A fast algorithm for computing prefix probabilities. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 57–69. https://doi.org/10.18653/v1/2023.acl-short.6

Omaki, A., Davidson White, I., Goro, T., Lidz, J., & Phillips, C. (2014). No fear of commitment: Children's incremental interpretation in english and japanese wh-questions [Publisher: Routledge _eprint: https://doi.org/10.1080/15475441.2013.844048]. *Language Learning and Development*, *10*(3), 206–233. https://doi.org/10.1080/15475441.2013.844048

Omaki, A., & Lidz, J. (2015). Linking parser development to acquisition of syntactic knowledge. *Language Acquisition*, *22*(2), 158–192.

O'Rourke, P. (2013). The interaction of different working memory mechanisms and sentence processing: A study of the p600.

Ovans, Z. L. (2022). *Developmental parsing and cognitive control* (Doctoral dissertation) [ISBN: 9798351468365]. University of Maryland, College Park. United States – Maryland. Retrieved July 25, 2023, from https://www.proquest.com/docview/2731081776/abstract/C81EB03A7C834A7FPQ/1

Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The two-body inversion effect [Publisher: SAGE Publications Inc]. *Psychological Science*, *28*(3), 369–379. https://doi.org/10.1177/0956797616685769

Pearl, L., & Sprouse, J. (2013). Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition*, *20*(1), 23–68.

Perkins, L., Feldman, N. H., & Lidz, J. (2022). The power of ignoring: Filtering input for argument structure acquisition. *Cognitive Science*, *46*(1), e13080.

Perkins, L., Knowlton, T., Williams, A., & Lidz, J. (2024). Thematic content, not number matching, drives syntactic bootstrapping. *Language Learning and Development*.

Pietroski, P. M. (2010). Concepts, meanings and truth: First nature, second nature and hard work [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0017.2010.01389.x]. *Mind & Language*, *25*(3), 247–278. https://doi.org/10.1111/j.1468-0017.2010.01389.x

Pinker, S. (1984). *Language learnability and language development*. The MIT Press/Bradford Books, Cambridge, Massachusetts.

Pinker, S. (1989). Learnability and cognition: The acquisition of argument structure. learning, development, and conceptual change series.

Pourtois, G., Peelen, M. V., Spinelli, L., Seeck, M., & Vuilleumier, P. (2007). Direct intracranial recording of body-selective responses in human extrastriate visual cortex. *Neuropsychologia*, *45*(11), 2621–2625.

Powell, L. J., & Carey, S. (2017). Executive function depletion in children and its impact on theory of mind. *Cognition*, *164*, 150–162.

Prevor, M. B., & Diamond, A. (2005). Color–object interference in young children: A stroop effect in children 31/2–61/2 years old. *Cognitive development*, *20*(2), 256–278.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism*. *Spatial Vision*, *3*(3), 179–197. https://doi.org/10.1163/156856888X00122

Quine, W. V. O. (1960). *Word and object*. MIT press.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of verbal learning and verbal behavior*, *22*(3), 358–374.

Rayner, K., Garrod, S., & Perfetti, C. A. (1992). Discourse influences during parsing are delayed. *Cognition*, *45*(2), 109–139.

Roark, B. (2001). Probabilistic top-down parsing and language modeling [Place: Cambridge, MA Publisher: MIT Press]. *Computational Linguistics*, *27*(2), 249–276. https://doi.org/10.1162/089120101750300526

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive psychology*, *38*(2), 259–290.

Schöppner, B., Sodian, B., & Pauen, S. (2006). Encoding action roles in meaningful social interaction in the first year of life. *Infancy*, *9*(3), 289–311.

Simon, T. J. (1997). Reconceptualizing the origins of number knowledge: A “non-numerical” account. *Cognitive development*, *12*(3), 349–372.

Simpson, G. B. (1984). Lexical ambiguity and its role in models of word recognition. *Psychological bulletin*, *96*(2), 316.

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.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological monographs: General and applied*, *74*(11), 1.

Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*(3), 227–267.

Stahl, A. E., Pareja, D., & Feigenson, L. (2023). Early understanding of ownership helps infants efficiently organize objects in memory. *Cognitive Development*, *65*, 101274.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643.

Sturt, P., & Lombardo, V. (2005). Processing coordinated structures: Incrementality and connectedness [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog0000_8]. *Cognitive Science*, *29*(2), 291–305. https://doi.org/10.1207/s15516709cog0000_8

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of verbal learning and verbal behavior*, *18*(4), 427–440.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension [Publisher: American Association for the Advancement of Science]. *Science*, *268*(5217), 1632–1634. Retrieved May 7, 2024, from https://www.jstor.org/stable/2888637

Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of memory and language*, *27*(6), 597–632.

Tatone, D., & Csibra, G. (2020). Infants infer different types of social relations from giving and taking actions. *CogSci*.

Tatone, D., Geraci, A., & Csibra, G. (2015). Giving and taking: Representational building blocks of active resource-transfer events in human infants. *Cognition*, *137*, 47–62.

Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of memory and language*, *39*(2), 195–217.

Trick, L. M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(2), 331.

Trueswell, J., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89–134.

Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In *Perspectives on sentence processing* (pp. 155–179). Psychology Press.

Tutunjian, D., & Boland, J. E. (2008). Do we need a distinction between arguments and adjuncts? evidence from psycholinguistic studies of comprehension. *Language and Linguistics Compass*, *2*(4), 631–646.

Ueno, A., & Matsuzawa, T. (2004). Food transfer between chimpanzee mothers and their infants. *Primates*, *45*, 231–239.

Uller, C., Carey, S., Huntley-Fenner, G., & Klatt, L. (1999). What representations might underlie infant numerical knowledge? *Cognitive development*, *14*(1), 1–36.

Valian, V. (1990). Logical and psychological constraints on the acquisition of syntax. In *Language processing and language acquisition* (pp. 119–145). Springer.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*(2), 105–143.

Waxman, S., & Gelman, R. (1986). Preschoolers' use of superordinate relations in classification and language. *Cognitive Development*, *1*(2), 139–156.

Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *Journal of experimental child psychology*, *99*(2), 75–95.

Weighall, A. R., & Altmann, G. T. (2011). The role of working memory and contextual constraints in children's processing of relative clauses. *Journal of child language*, *38*(3), 579–605.

Wellwood, A., Xiaoxue He, A., Lidz, J., & Williams, A. (2015). Participant structure in event perception: Towards the acquisition of implicitly 3-place predicates. *University of Pennsylvania Working Papers in Linguistics*, *21*(1), 32.

Wexler, K., & Culicover, P. W. (1980). Formal principles of language acquisition.

White, A. S., & Lidz, J. (2022). Lexicalization in the developing parser. *Glossa Psycholinguistics*, *1*(1).

Williams, A. (2015). *Arguments in syntax and semantics*. Cambridge University Press.

Winter, Y. (2018). Symmetric predicates and the semantics of reciprocal alternations. *Semantics and pragmatics*, *11*, 1–EA.

Woodard, K., Pozzan, L., & Trueswell, J. (2015). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of experimental child psychology*, *141*, 187–209. https://doi.org/10.1016/j.jecp.2015.08.005

Woodard, K., Pozzan, L., & Trueswell, J. C. (2016). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of experimental child psychology*, *141*, 187–209.

Ye, Z., & Zhou, X. (2009). Executive control in language processing. *Neuroscience & Biobehavioral Reviews*, *33*(8), 1168–1177.

Yu, X., Li, J., Zhu, H., Tian, X., & Lau, E. (2024). Electrophysiological hallmarks for event relations and event roles in working memory. *Frontiers in Neuroscience*, *17*, 1282869. https://doi.org/10.3389/fnins.2023.1282869

Yuan, S., & Fisher, C. (2009). “really? she blicked the baby?” two-year-olds learn combinatorial facts about verbs by listening. *Psychological science*, *20*(5), 619–626.

Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child development*, *83*(4), 1382–1399.

Zhou, P., Shi, J., & Zhan, L. (2021). Real-time comprehension of garden-path constructions by preschoolers: A mandarin perspective. *Applied Psycholinguistics*, *42*(1), 181–205.