

ABSTRACT

Title of Dissertation: ARGUMENT ROLES IN ADULT AND CHILD
COMPREHENSION

Lara Maria Isabel Ehrenhofer
Doctor of Philosophy in Linguistics
2018

Dissertation directed by: Professor Colin Phillips
Department of Linguistics

Language comprehension requires comprehenders to commit rapidly to interpretations based on incremental and occasionally misleading input. This is especially difficult in the case of argument roles, which may be more or less useful depending on whether comprehenders also have access to verb information. In children, a combination of subject-as-agent parsing biases and difficulty with revising initial misinterpretations may be the source of persistent misunderstandings of passives, in which subjects are not agents. My experimental investigation contrasted German five-year-olds' argument role assignment in passives in a task that combined act-out and eye-tracking measures. Manipulating the order of subject and voice (Exp. 4.1, 4.3) did not impact German learners' success in comprehending passives, but providing the cue to voice after the main verb (Exp. 4.2) led to a steep drop in

children's comprehension outcomes, suggesting that the inclusion of verb information impacts how young comprehenders process argument role information.

In adults, many studies have found that although argument role reversals create strong contrasts in offline cloze probability, they do not elicit N400 contrasts. This may be because in the absence of a main verb, the parser is unable to use argument role information. In an EEG experiment (Exp. 5.1), we used word order to manipulate the presence or absence of verb information, contrasting noun-noun-verb reversals (NNV; *which cowboy the bull had ridden*) with noun-verb-noun reversals (NVN; *which horse had raced the jockey*). We found an N400 contrast in NVN contexts, as predicted, but surprisingly, we also found an N400 contrast in NNV contexts. Unlike previous experimental materials, our stimuli were designed to elicit symmetrically strong and distinct verb predictions with both canonical and reversed argument role assignments. These data suggest that adult comprehenders are able to overcome the absence of a main verb when probability distributions over combined verb-argument role information can contribute to generating role-specific verb candidates.

The overall investigation suggests that prediction and comprehension of argument role information is impacted by the presence or absence of verb information, which may allow comprehenders to bridge the divide between linguistic representations and world knowledge in real-time processing.

ARGUMENT ROLES IN ADULT AND CHILD COMPREHENSION

by

Lara Maria Isabel Ehrenhofer

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy in Linguistics
2018

Advisory Committee:
Professor Colin Phillips, Chair
Dr Yi Ting Huang
Dr Ellen Lau
Dr Jeffrey Lidz
Dr Nan Bernstein Ratner, Dean's Representative

© Copyright by

Lara Maria Isabel Ehrenhofer

2018

Acknowledgements

I've spent almost five full years at the University of Maryland, College Park. Apart from the decade when I lived in Berlin, this is the longest I've been in one place in my entire life. My time here has taught me the importance of community, and so I'd like to take a moment to thank the numerous people who have contributed to my growth as a scholar and as a person during my PhD.

The Maryland Linguistics department, and the wider community of the Language Science Center, provides that rare environment where everyone, from fellow students to lab managers to faculty members, consider themselves not just colleagues, but also friends and mentors. The generosity of this community in offering their time and support, across both personal and academic contexts, is truly astounding. The greatest thanks are due to my primary advisor, **Colin Phillips**. In our weekly meetings and on many other occasions, his lessons have gone far beyond what I came to get a PhD in. He taught me to value my struggles in learning as an opportunity for growth as a researcher and as a person. My co-advisor, **Bill Idsardi**, supported me in expanding my repertoire of knowledge in phonological theory (which I regret is completely absent from this dissertation). **Yi Ting Huang** and **Jeff Lidz** have been pillars of support (and much-needed critique) in my child comprehension work, featured in Chapter 4 of this dissertation. That project could not have been carried out without the help of our colleagues **Kazuko Yatsushiro** (Leibniz-Zentrum für Allgemeine Sprachwissenschaft, Berlin, Germany), **Barbara Höhle** and **Tom Fritzsche** (Universität Potsdam, Germany). It took a herculean effort

to collect data from over 100 child participants, spread out over four daycare centres in Berlin and the Potsdam BabyLab, and I'm grateful for their good cheer in helping me making this happen in record time. Thanks also go to **Ellen Lau**, who made important contributions to the EEG experiment described in Chapter 5, and who is a model of thoughtful commentary and an incredible teacher. For five out of my six semesters as a TA, **Peggy Antonisse** was rock of support through some personally difficult times.

The boundary between friends and colleagues is blurry among fellow graduate students. **Allyson Ettinger** has been a loving and loyal friend, cohort-mate and housemate. **Phoebe Gaston**, **Anton Malko** and **Natalia Lapinskaya** have been great friends, and they have repeatedly saved my bacon when it gets to coding and stats. **Jeff Green**, **Zach Stone** and **Chris Neufeld** have been wonderful people to learn with and spend time with. I am also grateful for the patience and forbearance of the many gentle souls with whom I have shared an office during my time here: **Shota Momma**, **Zoe Schlueter**, **Eric Pelzl**, **Mina Hirzel** and **Tyler Knowlton**.

None of the work I present here could have been done without the hard work of a large number of research assistants and lab managers. They helped develop faster and cleverer ways to get things done, and often contributed substantially to the content and direction of these projects. Thanks are due to most notably to **Neomi Rao**, **Julia Buffinton** and **Anna Namyst**, but also (in no particular order) to **Maggie Kandel**, **Jon Burnsky**, **Kerianna Frederick**, **Bethany Dickerson**, **Ilia Kurenkov**, **Hanna Muller**, **Sade Evans**, **Caitlyn Roberts**, **Brooke Rothman**, **Scott Kaplowicz**, **Sofia Zenke**, **Marlena Jakobs**, **Maite Seidel**, and **Leonie Herz**.

Table of Contents

Acknowledgements.....	ii
Table of Contents.....	iv
List of Tables	ix
List of Figures.....	x
List of Abbreviations	xi
1 Introduction.....	1
1.1 Why this investigation?.....	1
1.2 Major debates in this dissertation.....	2
1.2.1 Commitment and prediction.....	2
1.2.2 World knowledge and linguistic representation.....	3
1.2.3 The time-course of argument role access and use.....	4
1.3 Outline of this dissertation	5
2 Argument role commitments in adult and child comprehenders.....	7
2.1 Introduction.....	7
2.2 Commitments and prediction	9
2.3 Evidence of comprehenders' commitments to argument roles.....	11
2.4 Pre-verbal argument role commitments.....	18
2.5 Conclusion.....	21
3 Argument roles in adult and child prediction	23

3.1	Introduction	23
3.2	Prediction from argument roles and verbs	24
3.3	Predicting verbs from arguments	28
3.4	Research questions	35
4	Argument role and verb commitments in German children's processing of passives	40
4.1	Introduction	40
4.2	Sources of variability in children's comprehension of passives	43
4.2.1	Subject-first biases and revision	45
4.2.2	Factors in passive comprehension	49
4.3	German passives: a preliminary corpus study	54
4.4	Experiment 4.1: Subject/Voice Manipulation	57
4.4.1	Participants	58
4.4.2	Materials	59
4.4.3	Procedure	62
4.4.4	Results	62
4.4.4.1	Behavioural data	64
4.4.4.2	Eye-tracking data	65
4.4.5	Discussion	70
4.5	Experiment 4.2: Verb/Voice manipulation	72
4.5.1	Materials	72
4.5.2	Participants	74
4.5.3	Procedure	74

4.5.4	Results	74
4.5.4.1	Behavioural data	74
4.5.4.2	Eye-tracking data	75
4.5.5	Discussion	79
4.6	Experiment 4.3: Control.....	81
4.6.1	Materials and Procedure	81
4.6.2	Participants	82
4.6.3	Results	82
4.6.3.1	Behavioural data	83
4.6.4	Discussion	84
4.7	General Discussion.....	85
4.7.1	Eye movements as an indication of argument role assignment.....	87
4.7.2	Alternative explanations: frequency, flexible word order, deficient linguistic representations, differences in experimental paradigms.....	90
4.7.3	Argument roles, verb information, parsing and commitments.....	95
4.7.4	Similar results in prior work.....	102
4.8	Conclusion.....	103
5	Prediction using pre-verbal and post-verbal argument role information	105
5.1	Introduction.....	105
5.2	“Semantic attraction”-based accounts of role reversal blindness	106
5.3	Prediction, competition and candidate generation	113
5.4	Argument roles and prediction mechanisms	116
5.5	Experiment 5.1	120

5.5.1	Materials	120
5.5.2	Procedure	123
5.5.3	EEG recording	124
5.5.4	Participants	125
5.5.5	Results	125
5.5.5.1	Behavioural results	125
5.5.5.2	EEG results	126
5.5.6	Discussion	127
5.5.7	Post-hoc stimulus analyses	128
5.5.7.1	Target cloze distribution	129
5.5.7.2	Entropy	130
5.5.7.3	Frequency	130
5.5.7.4	Subject-verb cosine relationship	131
5.5.7.5	Plausibility: a follow-up experiment	132
5.6	General discussion: Argument roles in prediction and retrieval	135
5.6.1	How could the format mismatch account explain the present results?	137
5.6.2	How could the probability-based account explain the present results?	141
5.6.3	Future directions	144
5.7	Conclusion	146
6	Argument roles in adult and child comprehension	148
6.1	Introduction	148
6.2	Verb information and argument role commitments	152
6.3	Structural commitments and prediction	159

6.4	Probabilistic factors in commitment and prediction	164
6.5	Timing in commitment and prediction.....	175
6.6	Parallels in adult and child argument role processing.....	178
6.7	Conclusion.....	180
7	Conclusion	182
	Appendix 1: Critical trial materials for Experiment 4.1	186
	Appendix 2: Critical trial materials for Experiment 4.2	188
	Appendix 3: Critical trial materials for Experiment 4.3	190
	Appendix 4: Critical stimuli for Experiment 5.1	192
	References.....	205

List of Tables

Table 4.1: Occurrences of <i>werden</i> in caretaker speech.....	56
Table 4.2: Experiment 4.1 sample stimuli	59
Table 4.3: Behavioural results, Exp. 4.1	65
Table 4.4: Average duration of regions of interest, Exp. 4.1	68
Table 4.5: Eye-movement statistics after auxiliary onset, Exp. 4.1	69
Table 4.6: Eye-movement statistics from pronoun onset, Exp. 4.1	70
Table 4.7: Experiment 4.2 sample stimuli	73
Table 4.8: Behavioural results, Exp. 4.2.....	75
Table 4.9: Average duration of regions of interest in Experiment 4.2	76
Table 4.10: Eye-movement analysis from verb onset, Exp. 4.2	77
Table 4.11: Eye movement statistics from auxiliary onset, Exp. 4.2.....	79
Table 4.12: Experiment 4.3 sample stimuli	82
Table 4.13: Behavioural results, Exp. 4.3	83
Table 5.1: EEG Sample Stimuli.....	122
Table 5.2: Mean plausibility ratings by test condition.....	126
Table 6.1: Mock distribution of nouns over argument roles with specific verbs (Ehrenhofer stimuli).....	170
Table 6.2: Mock distribution of nouns over argument roles with specific verbs (Chow et al., 2015 stimuli)	170

List of Figures

Figure 4.1: Eye movements after auxiliary onset (Exp. 4.1)	67
Figure 4.2: Eye movements after pronoun onset (Exp. 4.1)	67
Figure 4.3: Eye movements after verb onset (Exp. 4.2)	76
Figure 4.4: Eye movements from auxiliary onset (Exp. 4.2).....	78
Figure 5.1: Sample EEG waveforms, Exp. 5.1	127

List of Abbreviations

Abbreviation	Stands for...
acc.	accusative
dat.	dative
EEG	electroencephalogram
EOG	electro-oculogram
N	noun
NNV	noun-noun-verb
nom.	nominative
NP	noun phrase
NVN	noun-verb-noun
obj., O	object
pers	person
pl	plural
SD	standard deviation
SE	standard error of the mean
sg	singular
SOA	stimulus-onset asynchrony
subj., S	subject
V	verb
VP	verb phrase

1 Introduction

1.1 Why this investigation?

In language processing, one of the comprehender's most important tasks is to understand who is doing what to whom. This is crucial to understanding the events being described, who took part, what the grammatical relation between the participants was (was the dog the agent of biting, or the patient? Did it bite or get bitten?) and what the outcome was – arguably a primary aim of speech and comprehension. Yet evidence from sentence processing studies has consistently indicated that extracting argument role information from sentences in real time, building sentence parses, using them to predict upcoming information and interpreting the overall parse, is a challenging task for adult and child comprehenders alike. This dissertation tackles the general issue of how roles are used in language comprehension from two angles. The perspective from child comprehension illuminates the structural commitments that children are able to make on the basis of role information. Because children's language processing often reflects the earliest stages of adult processing (see Phillips & Ehrenhofer, 2015, for a review), investigating children's role commitments helps shed light on the starting-point for adult processing. The perspective from adult processing goes a few steps further, providing finer-grained detail on how adults succeed or fail when they generate predictions using the role commitments they have formed on the basis of an incomplete sentence parse. Investigating prediction and its shortcomings allows us to see more clearly how linguistic memory is accessed in real time, and make informed

conjectures about what the architecture of linguistic memory must be in order to give rise to the access patterns we observe. This in turn informs the debate over why children's early processing of argument role information faces the obstacles it does.

1.2 Major debates in this dissertation

Before I lay out how this investigation will unfold, I would like to provide a bird's-eye view of the major debates that my work touches on. While discussing each of them in full will not be possible within the limits of this dissertation, it is worth highlighting a few key contrasts that will thread themselves through the framework of this investigation.

1.2.1 Commitment and prediction

Written or spoken language unfolds as a flat linear sequence of morphemes, but language comprehension relies on imposing an ordered syntactic structure on this linear sequence, and then using it to access memory to generate expectations about upcoming sentence material, which will in turn lighten the load of making structural sense of incoming material. Committing to a structural parse of a sentence and using that parse to generate predications about upcoming material are therefore iterative, interdependent processes in language comprehension. No discussion of sentence processing should limit itself to a treatment of just one of these topics; I outline the relationship between them in more detail in section 2.2. However, the existing

experimental literature has tended to focus on either prediction or commitment, and it has often been split along methodological lines. For investigating child comprehension, the most suitable methods are behavioural tasks that can yield reliable results without requiring large participant sample sizes, many experimental trials, or extensive equipment set-up. These methods necessarily investigate the overall outcome of children's comprehension. They can be supplemented with time-sensitive methods such as eye-tracking, and often discrepancies between online and offline measures of comprehension show interesting cracks in the foundations of language processing (Lewis & Phillips, 2015). I contribute to this broad-based investigation of both the commitment and prediction aspects of language processing by using a combination of online and offline measures to investigate both commitment and prediction in adult and child comprehenders.

1.2.2 World knowledge and linguistic representation

Comprehenders use linguistic representations to access world knowledge. Argument roles interestingly bridge this divide, and there has long been vivid discussion of exactly where the boundaries lie. In particular, there is a question of how sensitively argument role information is encoded in verbs. Detailed discussion lies far beyond the scope of this dissertation. However, let it be noted that there appears to be widespread agreement that, in the words of Tanenhaus, Carlson, & Trueswell (1989), verbs “provide information that is used in resolving thematic ambiguity.” Part of the aim of this investigation, then, is to identify how dependent comprehenders are on waiting for the verb in using argument role information. According to a number of theorists

(see Williams, 2015 for a review), verbs provide the events to which roles relate. However, as we will see, comprehenders are able to extract argument role information even prior to encountering a verb. We examine whether there is a qualitative contrast in the argument role commitments that comprehenders are able to make with and without a verb, and what the outcomes are for processing.

1.2.3 The time-course of argument role access and use

Sentence comprehension is an incremental process, and this means that comprehenders have access to different information at different points in time, which plays a large role in how they are able to access linguistic memory. The incomplete evidence that comprehenders have also places them at risk for various kinds of processing error. This of course is a boon to the investigator, as the best way to learn how a complex system is not by observing it in its optimal function, but to carefully dismantle it when it makes errors: first to understand how it works, then to fix the mistake. In both of the experimental investigations in this dissertation, I use contrasts in the order of the information that comprehenders receive to illuminate how the availability of different types of linguistic representation influences comprehenders' language processing outcomes.

1.3 Outline of this dissertation

Chapters 2 and 3 are designed to provide the background information that is required to motivate the experimental investigations. In Chapter 2, I focus on the argument role commitments that comprehenders make in the presence or absence of verb information. A fairly large segment of explanations of why adult and child comprehenders experience difficulty in using argument role information argue that this is because argument role information is at best underspecified in the early stages of processing. I instead show that while the grain size of comprehenders' argument commitments may be fairly coarse, comprehenders nonetheless make these commitments and use them as the basis of further processing. This provides an important backdrop for Chapter 3, which explores how comprehenders use argument role commitments in prediction and outlines in greater detail the research questions for this dissertation. Chapter 4 describes the first experimental investigation: an exploration of German-speaking children's comprehension of passives. The flexible word order of German provides an excellent opportunity to contrast five-year-old's argument role commitments when they do or do not have access to verb information, with striking consequences. Chapter 5 contrasts adult comprehender's argument-role-based predictions when they do and do not have access to verbs in object-relative and subject-relative clauses. Contrary to prior literature, which has found repeatedly that adults experience considerable difficulty in predicting verbs from argument roles, we find that this result is more variable than previously reported. Chapter 6 works towards a unified explanation of the experimental phenomena found in this dissertation and elsewhere, proposing a view of argument roles as bridging the divide

between linguistic representations and world knowledge, with important ramifications in online sentence processing. Chapter 7 concludes.

2 Argument role commitments in adult and child comprehenders

2.1 Introduction

My larger investigation revolves around the question of how comprehenders identify who does what to whom in a sentence and how their ability to use this information in further language processing is affected by the presence or absence of verb information. To gain insight to these complex questions, we will need to understand the evidence that is available from two sides of comprehension: commitment and prediction. The literature has tended to treat these as separate fields of inquiry, and in section 2.2, I will lay out the separation I am pursuing in this chapter and in Chapter 3, while acknowledging that commitment and prediction are interrelated aspects of language processing.

In section 2.3, I discuss the extent to which comprehenders make structural commitments to argument roles, or to grammatical functions. In a sentence like *the rodeo clown knew which cowboy the bull had gored*, the presence of a second NP in the noun-noun-verb linear order of the embedded clause allows the comprehender to infer that the first NP is an object. Reviewing a relatively vast literature, there is ample evidence that both adult and child comprehenders commit quickly and firmly to interpreting incoming NPs in their grammatical functions, e.g. subject or object, using cues from word order. It is less clear to what extent a commitment to a specific syntactic position (e.g. *bull-OBJ*, *cowboy-SUBJ*) feeds forward into a commitment to a

corresponding argument role (e.g. *bull-THEME, cowboy-AGENT*). In part, this is due to methodological complications: it is difficult to design a sentence comprehension study that adequately dissociates syntactic position from argument role without running the risk of introducing other sources of processing difficulty, as this frequently entails using non-canonical and infrequent constructions like object-relative clauses or passives. For evidence regarding children's argument role commitments, I will examine the passive comprehension literature, which investigates a case of dissociation between syntactic positions and their canonical argument role assignments. Despite the methodological limitations of many of the earliest studies in particular, children's well-documented and robust difficulties with passive comprehension indicate that they, too, commit to subject-as-agent, object-as-patient parses.

Chapter 1 briefly discussed the question of what argument role information is encoded in verbs, and to what extent online processing is able to draw on this information before a verb has been encountered, as well as the balance between linguistic cues and other factors like world knowledge or thematic fit in comprehension. Section 2.3 addresses these questions by investigating the extent to which comprehenders are able to form argument role commitments prior to encountering a verb. This then sets the stage for the discussion in Chapter 3 of the extent to which adult and child comprehenders are able to use argument role commitments in predicting upcoming sentence material.

2.2 Commitments and prediction

It is beyond dispute that language comprehension involves both the formation of structural commitments, which impose a syntactic structure on the flat linear order of incoming sentence material, and prediction, which involves generating expectations about upcoming words, concepts, and structural elements. It is also unequivocally true that these are interrelated aspects of comprehension. In order to be able to form expectations of future lexical items, the comprehender needs to have built a structural parse of past material. At the same time, forming a structural parse can involve the prediction of upcoming structure, e.g. when long-distance dependencies are parsed. Contrary to appearances that some low-level forms of prediction are insensitive to structure, the cases of prediction under discussion in this dissertation clearly require past material (e.g. subjects and objects) to be integrated in a structural parse, and prediction of upcoming verbs is sensitive to that structure – though the precise extent of that sensitivity is the topic of debate in Chapters 3 and 5. For instance, Chow et al. (2015) demonstrate that the N400 differentiates between plausible and implausible arguments of a target verb in object-relative clauses. They contrasted sentences in which the two nouns were always equally highly associated with the target verb, but were either plausible or implausible participants in the event the verb described (plausible: *which tenant the landlord had evicted*, implausible: *which realtor the landlord had evicted*) and found higher N400 amplitudes in the implausible condition. This indicates that prediction is precise enough to generate verb candidates that are specific to a pair of preceding noun phrases (*evict* is not predicted as a plausible verb target of *realtor* and *landlord*). Nonetheless, in the same study, a different experiment

that reversed the order of arguments in object-relative clauses showed a lack of sensitivity to argument roles, with facilitated N400 amplitudes to target verbs in sentences like ...*which waitress the customer had served*. This indicates that prediction, while able to take into account the lexical content of both arguments of the to-be-predicted verb, does not reflect the roles that these arguments take, leading to a lack of differentiation between the verb predictions generated by *waitress-Agent, customer-Patient* and *customer-Agent, waitress-Patient*. In order to understand the limits of prediction in these cases, it is necessary to first understand how structural parses are initially built, and to what extent they incorporate information about argument roles. This is the aim of Section 2.3.

A further point to note prior to embarking on this discussion is that “commitment” has been used to refer not just to the initial structural parse a comprehender builds but also to the frequent persistence of such a parse, even in the presence of information which should alert the comprehender to the incorrectness of their structural interpretation. Especially in the child literature, discussion of comprehenders’ persistent “commitment” to particular parses has dominated over investigations of the nature of these initial commitments when they are formed. This is at least in part because the methodological limitations of working with children forced a focus on offline measures, which necessarily deal mostly with questions of why young comprehenders seem to fare so much worse in revising initially incorrect parses than their adult counterparts. This form of lingering commitment has been observed in adults, too (Christianson et al., 2001 and others), where it is taken as an indication that once a structural parse has been built and interpreted, reanalysis

creates a new proposition without erasing the previous one from memory (Slattery et al., 2013). Both meanings of “commitment” provide useful insights into how structural parses are built, and to what extent they are built to last. Nonetheless, the focus of the present chapter will be to illuminate the extent to which the initial parses children build differ from, and are similar to, those built by adults, with a view to exploring differences in their revision and persistence in Chapters 4 and 6.

2.3 Evidence of comprehenders’ commitments to argument roles

One of the thornier debates at stake here is the issue of whether comprehenders initially build structural parses that fully represent argument roles, or whether their structural commitments only reflect syntactic positions. As we will see in Chapters 3 and 5, using argument roles in prediction appears to be difficult for adult comprehenders. In this section I will show that adults have a tendency to incrementally interpret an initial NP as a subject, even when that initial NP is ambiguous as to whether it is a subject or object. I will then review comparable evidence from child processing studies. Finally, I will discuss to what extent this structural commitment to a subject-initial reading extends into a thematic subject-as-agent interpretation.

Languages with flexible word order provide an opportunity to investigate adults’ parsing of initial NPs. Dutch or German, for instance, disambiguate between subject- and object-relative clauses through case marking on nouns and in number on the verb in an NNV clause, rather than (as in English) making a distinction in verb

position between OSV and SVO word order in relative clauses. In Dutch and German, subject-relative clauses are generally read faster overall than object-relatives (Frazier, 1987; Hemforth & Konieczny, 2000). In more extensively case-marked languages like German, disambiguation between SO and OS orders can occur on the first noun, though this is complicated by the overall scarcity of object-initial clauses as well as widespread syncretism in the case paradigm, which in practice frequently makes it difficult to differentiate between initial objects and subjects. Bates, MacWhinney & Kliegl (1984) claim that case-marking is ambiguous in as much as 70% of all sentences. Corpus work by Matzke, Mai, Nager, Rüsseler, & Münte (2002) shows that where case-marking on an initial NP is ambiguous between subjects and objects, the initial NP is a subject in 93% of cases. Bader & Häussler's (2010) extensive corpus study shows that OS orders are rare (ca. 3%), but seem to be preferred if they allow nouns to be ordered by animacy within the sentence. Bornkessel, Schlesewsky & Friederici (2000) show that there are gradations within the dispreference for initial objects: although accusative marking is more frequent overall than dative marking, initial accusative objects yield distinctive ERP profiles over initial dative objects, mirroring the frequency statistics found by Bader & Häussler (2010). In languages that allow flexible word order and use case marking as an aid (if not always a fool-proof guide) to nouns' syntactic position as subject or object, there is nonetheless a subject-first parsing preference, and even when verbs are biased towards an object-relative reading, this preference is borne out (Schriefers, Friederici, & Kühn, 1995). Stacking the deck towards an object-initial parse by using inanimate NPs also does not result in a change in subject-initial parsing preferences in

Turkish (Demiral, Schlesewsky, & Bornkessel-Schlesewsky, 2008). The preference for a subject-initial reading also extends to wh-fillers in German: comprehenders assume that wh-gaps refer back to the subject (Schlesewsky, Fanselow, Kliegl, & Krems, 2000).

For children, as for adults, it is difficult to design studies that differentiate between parses that reflect children's commitments to syntactic positions versus argument roles. Children's patterns of non-adultlike language comprehension patterns show that they interpret initial NPs as subjects.

The first body of evidence comes from studies investigating children's comprehension of non-canonical word orders. A small set of studies shows that although children's sensitivity to case marking and word order develops by the age of 6, this sensitivity can be relatively slow to be reflected in behavioural measures. Schipke, Knoll, Friederici, & Oberecker's (2012) findings pattern closely with those of Matzke et al. (2002) for adult comprehenders. Schipke et al. (2012) found that at the age of 5, German-speaking children's ERP responses to unambiguously case-marked initial NPs do not differentiate between objects and subjects, although this contrast emerges by the age of 6. However, even at this age, children's behavioural responses follow a subject-initial preference pattern. Schipke, Friederici, & Oberecker (2011) demonstrated that while German children as young as three years show an adultlike P600 response to the second NP in illicit double-nominative (*SSV) constructions, even by the age of 6 there is no similar adultlike P600 response to the second NP in illicit double-accusative (*OOV) constructions. Taken together, these findings suggest that the developing parser is capable of noticing case-marking

violations, or unexpected or infrequent case assignment relative to canonical word order. However, as the critical stimuli used in this study were ungrammatical sentences, and the procedure did not include a behavioural task, it is difficult to draw conclusions about the nature of any processing children engaged in beyond noticing the case anomaly.

Several studies investigate the extent to which children are able to use case morphology by pitting it against other information, such as word order or animacy. Schaner-Wolles (1989) demonstrated that before the age of 5, German-speaking children are unable to use case-marking information if this conflicts with canonical word order (e.g. in an OVS construction), and that they instead prioritise word order information. This result was also found for novel verbs (Dittmar, Abbott-Smith, Lieven, & Tomasello, 2008). In a sentence imitation task, Diessel & Tomasello (2005) report that young learners of English as well as German tend to repeat object-initial sentences as subject-initial sentences up until age 5, indicating that the distinction between subject and object is erased in preparation for production at this age. In an act-out study comparing various types of complex relative clause construction, de Villiers, Flusberg, Hakuta, & Cohen (1979) found that children interpreted the object of an object-relative clause as its subject (i.e. the agent of the action). Lindner (2003) found that, when faced with unambiguously case-marked NPs, German-speaking children aged 2-3 rely on animacy cues to agenthood, but switch to using word order by age 4. As noted, these studies show children's parsing preferences in terms of syntactic positions, rather than in terms of argument roles. We see from this evidence that even in languages like German, where children are

exposed to more flexible word orders than e.g. English, and where additional cues in the form of case-marking are available, young comprehenders nonetheless tend to apply a one-size-fits-all parsing heuristic that frequently incorrectly assumes that the initial NP is a subject. But to what extent does this transfer into committing to a structural parse of a subject as an agent?

One place to investigate the extent of children's argument role commitments is to evaluate how they fare with sentences in which argument roles are assigned non-canonically. In passives, syntactic positions and argument roles do not align as usual: subjects are patients, whereas objects are agents. Assessing children's comprehension of passives therefore gives an indication to the extent of their expectation that a subject be an agent. By all accounts, child comprehenders encounter great difficulty in comprehending passives. Although early studies of child comprehension suffer from a variety of methodological issues, on the whole, behavioural data show that children's most common comprehension error in passives reflects a subject-as-agent misinterpretation. The first major breakthrough in the study of children's comprehension of passives came from the findings demonstrating that their failure to comprehend passives is not monolithic. Children fare better in comprehending passives with actional verbs (e.g. *hit*, *kick*) than non-actional verbs (e.g. *like*, *hate*; Gordon & Chafetz, 1990; Harris, 1976; Maratsos, Fox, Becker, & Chalkley, 1985), though their accuracy still lags behind comprehension accuracy in actives. This discrepancy can probably be at least partially traced to the non-imageability of non-actional verbs, which would make it difficult to accurately select a picture or unambiguously act out a sentence. Nonetheless, this finding hints that children's

difficulty with passives may be graded, rather than absolute, and subsequent studies have upheld this. Pinker, Lebeaux, & Frost (1987) extended the finding of a contrast between the comprehension of actional and non-actional passives to novel verbs. In their Experiment 4, they taught children novel verbs whose meanings required the subject to be a patient and the object an agent; despite their conclusion that children's lower probability of acting out a passive for these verbs indicates a selective resistance to passivisation on the basis of their knowledge of argument roles, it seems difficult to trace this effect beyond a general resistance to passive production in act-outs or speech. Again using novel verbs alongside familiar ones, Dittmar, Abbot-Smith, Lieven, & Tomasello (2014) find that German and English-speaking children do not reliably score above chance level in passive comprehension until the age of 4;7. Turner & Rommetveit (1967) found that children fare better at interpreting non-reversible passives, i.e. ones in which the reversed argument role assignment was implausible; although their materials contained animacy confounds, the same basic finding was replicated by Strohner & Nelson (1974). De Villiers & de Villiers (1983) found a correlation between children's mean length of utterance and their passive comprehension accuracy, although their likelihood of displaying a role-reversed interpretation also rose as their mean length of utterance expanded. The authors suggest that as children's ability to produce syntactically complex structures improves, their parsing strategy initially comes to rely on a subject-first parsing heuristic.

So far, the main take-away from these studies has been that children appear to form structural parses in which they commit to interpreting subjects as agents, and

this remains fundamentally the case until roughly the age of 5-6. Until this point, various factors may influence children's likelihood of correctly interpreting a passive, such as the probability of the argument role assignment in the overall event. In terms of an explanation of why children's comprehension of passives remains poor for so long, many scholars have cited frequency (e.g. Demuth, 1990) or an underlying lack of representational capacity to represent passives (e.g. Borer & Wexler, 1987; Fox & Grodzinsky, 1998); these accounts are discussed in more detail in Chapter 4. However, in combination with the extensive literature on children's difficulties with revision (Choi & Trueswell, 2010; Engelhardt, 2014; Trueswell, Sekerina, Hill, & Logrip, 1999; Weighall, 2008 and many others), recent work (Huang, Leech, & Rowe, 2017; Huang, Zheng, Meng, & Snedeker, 2013) has suggested that the root of children's difficulty with passive comprehension lies in their initial argument role commitments, in which they interpret subjects as agents but are then unable to revise this commitment.

If this is true, creating situations where argument role assignments are easier to revise should lead to a higher success rate in passive comprehension. Huang et al. (2013, 2017) tested comprehension in English and Mandarin, contrasting children's comprehension of passive sentences in which the subject was a full NP (*the seal was quickly eaten by it*) or a pronoun (*it was quickly eaten by the seal*). Across both studies, the contrast in children's behavioural responses indicated greater success in revising argument role assignment when the subject was a pronoun.¹ This suggests that there is some gradience in children's argument role commitments, or their

¹ Note that in the Mandarin study, this contrast was in fact largely driven by a reduction in children's performance in actives, rather than an actual improvement of their performance in passives.

capacity to revise an incorrect commitment: if a pronoun cannot be resolved until the full sentence information (including the verb and the other event participant) are known, children are unable to definitively identify the agent until late in the sentence. To be clear (and in slight deviation from the discussion in Huang et al. 2013, 2017), this does not mean that children are not committing to subjects as agents. It does mean, however, that while the referent of a subject-agent pronoun is unknown, children are able to reassign argument roles in a manner that ultimately allows them to be more successful in comprehending passives. This will be discussed in more detail in Chapters 4 and 6.

In this section, I examined evidence from several intersecting literatures to investigate the extent to which adult and child comprehenders' structural parses contain commitments to argument roles. In large part, this focused on a discussion of the extent of comprehenders' preference to interpret initial NPs as subjects, which in the case of both adults and children also entails a commitment to the subject as an agent. The next part of this discussion will revolve around the question of whether there is a qualitative difference between the argument role commitments that adult and child comprehenders make when they do or do not have access to verb information.

2.4 Pre-verbal argument role commitments

The previous section described a wealth of literature that applied the tools of psycholinguistics to well-known syntactic phenomena like gap sites and wh-movement. However, attempting to refract that focus through the lens of the

commitments that the parser can make based on preverbal information is less illuminating, for two reasons. Firstly, although verb-final or partially verb-final languages like Japanese or German are fairly well-studied, the output of studies on these constructions is nonetheless far smaller than the number of available studies on canonical SVO constructions in English. Secondly, within the relevant literature, experimental work has been focused on different debates that do not easily align with the questions at stake here. For instance, experimental interests have frequently been turned onto the investigation of attachment preferences or subject-object ordering preferences (e.g. Bader & Meng, 1999; Schriefers et al., 1995), but this work again frequently focuses on questions of structural preferences e.g. whether initial NPs are preferentially interpreted as subjects, or using evidence from garden-path effects in response to PP attachment ambiguities to shed light on the debate over whether the processor constructs parses serially or in parallel, in two steps or in one.

A salient debate in this corner of the literature has been whether the parser can initiate any kind of structure-building when the verb head has not yet been encountered. Head-driven theories of parsing (see Kamide & Mitchell, 1999 and references therein) predict that this should not be the case, and consequently a body of studies from the 1990s showed that this prediction is not consistent with extant data. The terms of the debate have focused more on the *whether* of preverbal commitments, not so much on the *how* or *what*, but this section will attempt to provide a small selection of pertinent findings.

Felser, Clahsen, & Münte (2003) investigated double object constructions in subordinate clauses in German, using a lengthy adjunct phrase to temporally separate

the initial object from the embedded subject. They found a LAN effect at the embedded subject and interpreted this as evidence of a working memory cost, suggesting that this cost is incurred by the parser's need to start constructing a VP at the embedded subject (in preparation for an actual V head). While this study provides some evidence that the parser is committing to a structural parse, it is less clear what the nature of that parse is, including to what extent it includes commitments on argument roles.

Kamide & Mitchell (1999) used a self-paced reading task to investigate low and high attachment preferences in Japanese, contrasting sentences in which the thematic properties of the final verb forced either a low or high attachment reading. Compared to a globally ambiguous baseline, they found elevated reading times at the matrix verb in the condition that only allowed the low attachment. This suggests that upon encountering a dative NP, the parser commits to high attachment structure, allowing for the prediction of a verb with specific thematic properties. This study therefore expands on the previous studies' findings: it confirms that the parser is making certain structural commitments, but also that these commitments are used as the basis for certain expectations concerning argument roles.

Christianson (2002) used an auditory version of a self-paced reading task to investigate sentence processing in Odawa (Algonquian), which provides a remarkable range of possible word orders in matrix clauses. Based on frequency, final verbs are predominantly of the "direct" category. (According to Christianson, Algonquian syntax is not yet fully understood, but direct verbs assign thematic roles canonically to preceding nouns, whereas inverse verbs flip the order of the thematic roles that are

assigned to preceding subject and object positions.) Speakers' listening times were higher when encountering a final direct verb than inverse verb, despite the fact that direct verbs assign argument roles canonically, unlike inverse verbs, which trigger a thematic role reassignment that is expected to cause processing delays. (There were no effects of verb type in spillover regions following the verb.) This result is unexpected, given previous findings that the parser makes certain structural commitments prior to encountering a verb.

In the realm of child processing, there is some evidence that children, like adults, commit to certain structural interpretations prior to encountering a verb. Unlike adults, however, they are unable to revise these commitments when they turn out to be incorrect. Choi et al. (2010) tested children's comprehension in Korean NNV sentences in which the initial noun's status as goal or location of the verb is ambiguous until revealed on the final verb. Even at the final verb, where the meaning and thematic roles made available by the verb should have allowed participants to disambiguate the two possible interpretations of the initial NP, children's eye gazes did not differentiate between goal and location interpretations. This result demonstrates that the developing parser, like the adult parser, makes structural commitments early, prior to encountering a verb.

2.5 Conclusion

The dilemma I outlined in Chapter 1 is two-fold: according to linguistic theory, argument role information is encoded on, and assigned by, main verbs. In processing, this raises questions about the extent to which this information can be used by

comprehenders before they have access to verb information. Sentence processing research has demonstrated that comprehenders are highly skilled at exploiting statistical regularities, so it is puzzling that despite the systematic correlations between argument roles and syntactic positions (e.g. subjects are frequently agents), comprehenders have difficulty exploiting argument role information in prediction. The aim of this chapter was therefore to explore how argument role information is incorporated in comprehenders' initial structural parses of incoming sentence material, and this question will inform the further discussion of how such information can be used in prediction in the next chapter.

I provided an overview of a range of sources of evidence showing that both adult and child comprehenders form commitments to syntactic interpretations of NPs as subjects or objects, and that there is a widespread cross-linguistic preference to assume that initial NPs are subjects. Much of this evidence comes from the processing difficulties that surface when it turns out that a sentence is non-canonical in its word order. Beyond this, the evidence as to whether comprehenders commit to argument roles when they commit to syntactic positions is difficult to evaluate. This will be in the background of discussion throughout the remainder of the dissertation, where we will be investigating the potential contrasts in commitments involving argument role info with or without verb info, and the extent to which this is useful in prediction.

3 Argument roles in adult and child prediction

3.1 Introduction

Chapter 2 laid out evidence that both child and adult comprehenders parse nouns in sentences according to their syntactic positions, but that over and beyond this, to an extent they also form commitments about the argument roles these nouns take. This is the case even when comprehenders have not yet encountered a verb, suggesting that the fine-grained argument role information contained in verbs is not necessarily required in order for comprehenders to eagerly posit argument role commitments in comprehension. The focus of this chapter will be on exploring comprehenders' ability to use these argument role commitments in online prediction. As before, the discussion will fall largely around the questions of what adult and child comprehenders are able to use argument role information for in prediction when they do or do not also have access to verb information.

Much of the literature I will discuss here ultimately revolves around questions of whether comprehenders use all bottom-up linguistic information as soon as it becomes available, and to what extent top-down forms of information fill in the gaps left by comprehenders' initial gaps in making use of available information. Some of the evidence laid out in Chapter 2 suggests that argument role information is represented even in young comprehenders' structural parses of incoming information. Part of this literature also engages in a longstanding discussion about how argument role information is represented: whether information like goodness of fit to argument

roles is represented lexically or whether this is retrieved from world knowledge, and if so, whether this occurs at a delay (see discussion in McRae & Matsuki, 2009).

This background will begin with an exploration of the extent to which comprehenders are able to use argument role information predictively when they have access to verb information (Section 3.1). Much of the evidence here comes from visual world eye-tracking studies, and part of the discussion will concern the difference between generating predictions in constrained or unconstrained contexts. The eye-tracking evidence initially indicates that adults are able to use argument role information easily in prediction when verb information is provided. However, even when verb information is provided and prediction takes place in a constrained context, the evidence suggests that prediction from argument role information is difficult for adult comprehenders. This will become even more apparent in Section 2, where I will move on to a discussion of how comprehenders fare in using argument role information in prediction when no verb is present. This will largely involve evidence from EEG studies showing adults' prediction in unconstrained contexts. In Section 3, I will lay out the research questions I hope to address in this dissertation (Section 3).

3.2 Prediction from argument roles and verbs

The first question to answer is about the extent of comprehenders' ability to use argument role information in prediction. The evidence that has been gathered around this issue mostly stems from two debates: one about the grain size of the argument

role representations that comprehenders use, and one about whether comprehenders are making predictions at the event level.

Evidence from eye-movement studies initially indicates that comprehenders are able to use argument role information to make predictions in a constrained context. Altmann & Kamide (1999) observed English-speaking comprehenders' eye movements in response to stimuli in which the verb either restricted the upcoming material (*the boy will eat the...*) or not (*the boy will move the...*). Comprehenders' eye movements indicated that they were able to rapidly use selection restrictions to identify the most likely upcoming referent in a field of competitors including a *cake* and a *train*. In the "eat" condition, comprehenders looked at the cake; in the "move" condition, their gazes were split between the two objects. This was the first study to show that comprehenders are able to use verb information to immediately narrow down the choice of referents in a visual array. In a follow-up, Kamide, Altmann, & Haywood (2003) expanded the study to investigate how English-speaking comprehenders combine argument and verb information in prediction. They found that the preceding noun influenced comprehenders' gaze towards the most likely second event participant in a scene showing a man, a girl, a motorcycle and a carousel: in the sentence context *The man will ride the...*, comprehenders' gazes oriented towards the motorcycle, whereas this preference flipped in the sentence context *The girl will ride the...* This shows that the selection restrictions of the verb alone are insufficient to determine predictive eye gazes, since both *motorcycle* and *carousel* objects fulfil the selectional restrictions of *ride*. Instead, comprehenders are forming predictions on the basis of nuanced information which takes the nature of the

agent into account. Knoeferle, Crocker, Scheepers, & Pickering (2005) showed that German-speaking comprehenders are able to rapidly extract argument role information from case marking in SVO or OVS sentences, even when case marking on the initial NP was ambiguous, to identify the correct second event participant.

So far, the evidence suggests that adult comprehenders are able to use verb information to predict upcoming sentence material in constrained contexts. However, there is some evidence to suggest that this language comprehension task is more difficult than meets the eye when it comes to argument roles. Kamide, Scheepers, & Altmann, (2003, Exp. 2) compared comprehenders' gazes towards target and competitor items in active and passive sentences (*The hare will eat/be eaten by the cabbage/fox*). They found that comprehenders were more likely to look at "fox" objects in the passive than in the active condition, which the authors interpret as evidence that comprehenders are able to quickly extract voice information from the verb to select appropriate targets in a visual world paradigm. However, in the passive condition, participants were numerically similarly likely to look at either "cabbage" or "fox" objects. This is similar to findings by Kukona, Fang, Aicher, Chen, & Magnuson (2011), who used a visual-world paradigm whose most relevant condition contained both plausible agents and patients of the verb, which itself was either predictive of the arguments or not. For instance, in the active version of the task, both a crook (target) and a cop (distractor) were present in a visual display while participants heard a stimulus like *Toby arrested/noticed the ...* In both active and passive versions of the task, the authors found that comprehenders did not initially differentiate between looks to the target and distractor. The findings discussed above

show that comprehenders can use lexical information from the verb (e.g. selectional restrictions) to form fine-grained predictions of upcoming verb arguments. However, these results from studies of argument selection suggest that when the competitor is a plausible argument of the predictive verb, comprehenders have greater difficulty correctly identifying targets.

Prediction from argument role information is a relatively underexplored segment of child comprehension. Although much has been made of children's ability to use information like grammatical gender to select between competing objects in a forced-choice selection task (Lew-Williams & Fernald, 2007, 2010), investigations of children's use of argument role information in prediction have been limited. Mani & Huettig (2012) replicated the *boy eats/moves the cake* paradigm with 2-year-olds, and found that children's gazes to the correct object in the display correlated with their vocabulary size, indicating that children are able to extract selectional restriction information from verbs in a manner similar to adults. Nation, Marshall, & Altmann (2003), using a highly similar paradigm with 10- and 11-year-olds, found that at older ages, reading comprehension did not predict children's overall probability of fixating the correct target, but did influence the duration of target fixations, indicating that prediction may be affected by working memory limitations.

Overall, the evidence from studies of prediction using argument roles and verb information suggests that both adult and child comprehenders are able to use a combination of subject and verb information to predict upcoming event participants. However, comprehenders' ability to do so is somewhat limited. Both adults and children are able to predictively use lexical information from verbs (e.g. selectional

restrictions). However, when adults are tasked with using argument role information from the verb to differentiate between two objects which are both plausible event participants, they do so with difficulty even in constrained choice paradigms (Kukona et al., 2011).

3.3 Predicting verbs from arguments

In Chapter 2, I reviewed evidence that adult comprehenders form commitments to argument roles even when they do not yet have access to verb information. This section will reveal to what extent comprehenders can use this information in prediction, especially in prediction of a verb.

Studies of prediction in constrained contexts show that adult comprehenders are able to use information about syntactic positions to predict upcoming sentence material when they do not yet have access to the verb. In a visual-world study in Japanese, Kamide, Altmann & Haywood (2003) tested sentences in which case marking on the second noun phrase revealed whether the sentence contained a direct object (*the waitress-NOM the customer-ACC merrily teased* vs. *the waitress-NOM the customer-DAT merrily the hamburger-ACC brought*), and found that comprehenders were able to selectively use case-marking to generate anticipatory eye movements towards the hamburger. In German, Kamide, Scheepers & Altmann (2003) tested comprehenders' looks towards likely themes or agents in OSV or SOV sentences with unambiguously case-marked nouns (*the hare-ACC will soon the fox-NOM eat*, “the fox will soon eat the hare”, vs. *the hare-NOM will soon the cabbage-ACC eat*, “the hare

will soon eat the cabbage”). They found that comprehenders reliably fixated on the correct agent (fox) in object-initial sentences, indicating that they had accurately extracted argument role information and used it to predict a likely upcoming target. However, proportions of looks towards likely themes (cabbage) did not vary by case marking, indicating that comprehenders had difficulty overcoming a subject-as-agent misinterpretation.

In these highly constrained contexts, it appears that comprehenders are able to use argument role information predictively, even prior to receiving verb information that they would be able to combine it with. However, to what extent does this carry over into the domain of freely generating predictions? Note that while the term “prediction” has been used in the literature to describe both forms of anticipatory processing, this collapses across two probably quite different processes: a multiple-choice task, in which comprehenders must select the most likely candidate from a finite set, contrasted with an essay question, which requires comprehenders to generate candidates from their entire lexicon. The available evidence suggests that once supportive visual information is removed, comprehenders have markedly more difficulty using argument role information. Role-reversed sentence contexts (in which both arguments are plausible event participants, but are assigned incorrect argument roles) typically engender no contrast in N400 amplitudes, though they are often accompanied by a P600 contrast. This general finding holds even with increasingly tightly controlled stimuli.

Comprehenders are able to differentiate between plausible and implausible arguments of the to-be-predicted verb. Chow et al. (2015, Experiment 4.2) tested

object-initial embedded indirect wh-questions in which the two noun phrases either were or were not plausible arguments of the target verb (... *which tenant/realtor the landlord evicted...*), and found that an N400 contrast here too, again indicating that prediction is sensitive to the distinction between arguments and non-arguments of a verb. Related results were found by van Herten, Kolk, & Chwilla (2005), who paired associated arguments with implausible verbs in Dutch SOV sentences (... *that the elephants-NOM the trees-ACC pruned/caressed...*) and again found an N400 contrast on the target verb, accompanied by a P600 contrast. Prediction is therefore sensitive to the distinction between arguments and non-arguments of the verb, and this distinction may be made on the basis of selectional restrictions or the combination of the plausibility of argument and verb information. This aligns with the studies discussed in the previous section: when comprehenders' task is to differentiate between plausible or implausible arguments of the verb, they succeed.

However, the findings paint a less clear picture when both noun phrases are plausible event participants of the verb, and the overall acceptability of the event depends on argument role assignment. Studies that varied the animacy of event participants showed that although animacy violations generate a P600 contrast, they do not yield an N400 contrast. Kuperberg et al. (2003) tested ERPs in sentences like *For breakfast the eggs/boys would only eat...* and found that the inclusion of an inanimate agent (*eggs*) did not increase N400 amplitude. The authors took the N400's blindness to the anomaly of an inanimate agent for the verb *eat* to indicate that comprehenders have greater difficulty taking argument role information into account when the noun is a plausible event participant. Chow & Phillips (2013) conducted a

set of studies in Mandarin which contrasted SOV sentences in which the final verb matched or mismatched the preceding context according to an animacy violation (e.g. *the student baffled the math problem*) or a “combinability” violation (*the student hanged the math problem*), and found that prediction (as measured through an N400 contrast) again was not affected by animacy, though violations entailed a P600 contrast. Kim & Osterhout (2005, replicated in Kim & Sikos, 2011) contrasted active and passive sentences in which the verb target revealed an argument role reversal (*The hearty meal was devoured/devouring...*) and again found no N400 contrast, though this manipulation, too, yielded a P600 contrast. These findings generated a long line of theoretical explanations blaming the putative existence of a separate semantic analyser for suppressing bottom-up syntactic information that would have resulted in a “semantic violation” that could result in an N400 contrast.

A further frontier of the role reversal literature is the careful control of offline cloze predictions. N400 amplitude has been repeatedly found to correlate with the offline probability of sentence completions (Kutas & Hillyard, 1984). One possible reason for why the N400 appears to be insensitive to argument role reversals is the difficulty of balancing stimuli such that both orders generate distinct completions. Hoeks et al. (2004) tested Dutch sentences in which the passive verb target (*the javelins were by the athletes thrown*) is highly predicted offline, but there were no similarly high-probability completions for actives (what do javelins do to athletes?). In fact, a post-hoc rating study showed that participants found it much more difficult to generate completions for sentence fragments in the reversed conditions. The authors found no N400 contrast on the verb target, but speculate that the contrast in

generating a final verb completion may be partially to blame for these results. Interestingly, however, when predictability of the final item is closely controlled for, the N400 nonetheless does not show sensitivity to argument role reversals (Chow et al. 2015).

Argument role reversal effects do appear to surface at a delay. Chow et al. (2018) compared canonical and role-reversed Mandarin SOV sentences that did or did not include an adverbial phrase before the final verb (reversal: *xiaotou ba jingcha (zai shangxingqi) zhua-le* “the thief arrested the cop (last week)”). The adverbial phrase provided a temporal buffer between the two nouns and the to-be-predicted verb. When comprehenders were provided with this extra time for processing, the N400 reflected a contrast between the reversed and canonical sentence conditions; however, this contrast disappeared when the adverbial was not present and comprehenders therefore saw the target verb immediately after the second noun phrase. These findings suggest that the problem with using argument role information in prediction is not so much one of extracting or representing this information, but one of using it accurately to predict upcoming sentence material. Understanding what determines how comprehenders use argument role information therefore provides important insights to how linguistic information is organised in memory.

The evidence we have reviewed so far shows that even with increasingly tightly controlled stimuli – taking into account factors like the animacy of the two nouns, the overall plausibility of the event, or the likelihood of offline continuations, to name just a few – it is very difficult for adult comprehenders to use argument role information in prediction. (The evidence on child prediction using argument roles

without verb information is scarce if not nonexistent, but if the typical patterns of adult and child comprehension hold, one would expect a language comprehension task that is challenging for adults to be nearly impossible for children.) The pattern of N400 insensitivity, but P600 sensitivity, to argument role reversals has led into two separate but related debates, both of which will be discussed in more detail in Chapter 5. The first is the question of the functional significance of the N400: initial accounts of its insensitivity to argument role information were based on the assumption that the N400 is an index of semantic integration. Moving towards a prediction-based account of the N400 (see, among others, Van Petten & Luka, 2012 for a review) opens up the possibility of investigating why it is that argument role information seems not to have an immediate impact on the amplitude of the N400. The second is the question of why exactly argument role information is not reflected in N400 outcomes. These, too, will be discussed in greater detail in Chapter 5, but it is worth briefly outlining the main accounts prior to laying out the broader aims of this dissertation.

Accounts of the N400's insensitivity to argument role information have rested on four main options (each listed with one representative publication in which this claim is put forward):

- i. Argument role information cannot initially be used in language processing because it is ignored at the earliest stages of sentence comprehension. (Ferreira, 2003)
- ii. Syntax and semantics are processed separately. When the output of these processing streams clashes, top-down information following considerations of semantic plausibility wins out over bottom-up information indicating the

argument roles of nouns in the sentence under consideration. (Kim & Osterhout, 2005)

- iii. Argument role information is represented in comprehenders' earliest parses of incoming sentence material. However, it is difficult to use in predicting specific lexical items, particularly when these are verbs, due to a mismatch between the format of the search probe (which is in the *NP-Role* format, e.g. *waitress-Agent*) and the events to be targeted in memory. (Chow et al., 2016)
- iv. Argument role information is represented in initial parses and can be used in prediction, but prediction is governed by probabilistic associations between lexical items and contexts, so it is this that determines whether argument role information impacts prediction. (Kuperberg, 2016)

While further investigation of these claims must wait until Chapter 5 to be discussed in greater detail, one fundamental tension is the issue of whether argument role information is represented at the earliest stages of comprehenders' parses of incoming sentence material. As argument role reversals generate P600 contrasts under the same conditions in which they do not generate N400 contrasts, it seems difficult to argue that comprehenders initially build rough representations of incoming material that do not take argument role information into account at all (claim i). Furthermore, given the speed and accuracy with which even young comprehenders are able to deploy argument role information when they also have access to a main verb, it is difficult to justify a claim that argument role information is selectively overridden by top-down plausibility considerations (claim ii). Claims iii and iv each attempt to provide a principled reason as to why argument role information does not appear to affect

comprehenders' predictions at early stages of processing. This may be due to a difficulty in matching the format of a noun that has been interpreted as having a specific argument role (e.g. *seal-Agent*) to the format of events in memory (claim iii). Alternatively, comprehenders' initial insensitivity to argument role information may be due to a delicate balance of probabilities in memory, such that argument role information is sometimes but not always recruited in prediction (claim iv).

3.4 Research questions

This section will survey the broad brushstrokes I laid out in Chapters 2 and 3, assessing the evidence concerning comprehenders' ability to extract argument role information in online comprehension and use it in prediction, to formulate the research questions for this dissertation.

In Chapter 2, we saw that in online comprehension, both adult and child comprehenders form structural commitments concerning syntactic positions, e.g. the contrast between subject and object. To an extent, these include argument role information. The evidence here was difficult to gather, due to the difficulty of dissociating between subject position and agent role, for instance. Nonetheless, we saw evidence showing the strength of both adult and child comprehenders' subject-first biases, coming from an exploration of garden-path effects in verb continuation biases, SO/OS ambiguity resolution, and, especially in children, their well-documented difficulty with passives. Young comprehenders' errors in passive

comprehension suggested a specific difficulty in identifying the subject as the patient, not agent, of the event.

We saw that regardless of whether comprehenders have access to verb information or not, they form commitments to argument role assignments. These have the capacity to impact comprehenders' commitments about upcoming structure, as well. There were hints of a possible qualitative contrast between the commitments that comprehenders are able to make depending on whether they have access to verb information or not. The gain size of these commitments is not entirely clear: for instance, it is unclear whether pre-verbal commitments differentiate between agent and experiencer roles.

In Chapter 3, we saw that using argument role commitments in prediction is not straightforward. An initial look at how comprehenders use argument role information in combination with verbs to predict upcoming event participants suggests that they do so effortlessly and accurately. This held even for very young comprehenders. However, the evidence for this conclusion came from an experimental paradigm in which participants' ability to predict upcoming sentence material came from the proportions of looks directed towards different objects in a constrained environment. Even in this forced-choice prediction paradigm, argument role information provided surprisingly difficult to use if the visual world contained competing plausible participants in the event described by the verb (Kukona et al., 2011). Moving into free-generation prediction paradigms, there was some evidence of difficulty in adult comprehenders' ability to use argument role information predictively when verb information was present, and this difficulty was revealed to be

profound once verb information was not available. The methodologies that can be used to probe prediction in unconstrained contexts are difficult to use with child participants, so there is extremely little evidence to suggest how children fare in prediction from argument roles when no verb information is present. However, given past patterns in child and adult comprehension, anything that is difficult for the mature comprehender is even more challenging for the developing parser.

We also saw, in both Chapters 2 and 3, that there is a tendency to explain comprehenders' failures in the use of argument role information in comprehenders as an indication that this information is not represented at whatever stage of processing is being probed ("good enough" accounts of parsing and prediction). Given the evidence that comprehenders can and do form argument role commitments in parsing, it seems that the debate deserves to be refocused. Evidence of comprehenders' difficulty in using argument role information in prediction needs to shift away from discussion of whether this information is taken into account at all in the earliest stages of online processing, and instead move towards a discussion of why comprehenders appear to have such difficulty in using this information in prediction. The experimental investigations in the rest of this dissertation are designed to address this question.

The pattern we have seen so far is that both adult and child comprehenders form argument role commitments, but for young comprehenders in particular, these commitments are difficult to revoke, whereas for adult comprehenders, they can be difficult to use in prediction. We saw an intersecting pattern of evidence where there may be a qualitative contrast between argument role information when it is and is not

tethered to verb information, both in terms of the commitments comprehenders make and in terms of the usefulness of that information in prediction. This leads into two sets of research questions, which will be addressed in the experimental investigations in Chapters 4 and 5.

For child comprehenders, what are the factors determining the strength of their commitments? Understanding this provides us with a better window into the time course of how young comprehenders extract and use argument role information in comprehension. In Chapter 4, I present research that investigates this by testing the time-course of children's subject-as-agent commitments, and how the nature of these commitments (including or excluding verb information) impacts children's ability to revise. In three act-out experiments that were combined with a visual world paradigm, German-speaking children heard sentences in which they received a cue to sentence voice (i.e. argument role assignment) either immediately before or after the subject (Experiments 4.1 and 4.3), or after the main verb (Experiment 4.2). The variation in children's success at revising subject-as-agent misinterpretations provided insights to contrasts in the nature of the argument role commitments they had formed with or without verb information. Children's language comprehension differs from adults' mostly in terms of the timing of processing and in terms of comprehenders' ability to revise initially incorrect parses (Phillips & Ehrenhofer, 2015). For this reason, investigating the factors affecting children's argument role commitments can provide important insights into the same processes in adults.

For adult comprehenders, how does the availability or absence of verb information impact the ability to use argument role information in prediction? Some

of the contrast in prediction outcomes that I discussed in this chapter may well be due to differences in experimental methodologies. However, it is also possible that these differences amplify an existing underlying qualitative contrast between argument role information that is or is not combined with verb information. Chapter 5 uses EEG methodology to contrast adult comprehenders' predictions using argument role information in object-relative (NNV) against subject-relative (NVN) clauses. Taken together, the experimental investigations in the following chapters will illuminate the extraction and use of argument role information in real-time language comprehension.

4 Argument role and verb commitments in German children's processing of passives

4.1 Introduction

In sentence comprehension, understanding who did what to whom is crucial: if *the shark ate the seal*, we need to know which of the two antagonists survived in this particular instance in order to follow the discourse. In canonical sentences in accusative languages, the structural cue of who is the subject (*the shark*) and who is the object (*the seal*) probabilistically aligns with the thematic role assignment of agent and patient. However, this alignment does not extend to passive sentences: if *the seal was eaten by the shark*, *the seal* is the subject, but also the patient of the sentence. This misalignment poses a significant challenge to online sentence comprehension, especially to child comprehenders, who frequently misinterpret the subject of a passive as the agent of the event labelled by the verb. In this study, we investigate the source of this subject-as-agent misinterpretation in children's comprehension of passives, as well as mitigating factors in avoiding or rescinding it. In particular, we investigate whether children's argument role commitments differ qualitatively depending on whether they also have access to a main verb, and whether this has an impact on child comprehenders' ability to revise out of an initial misinterpretation. The flexible word order of German, our test language, provides an opportunity to isolate the contribution of different cues to children's processing of passives.

Children's difficulty with comprehending passives is well-documented. While accounts differ as to the exact reasons why passives are challenging for young comprehenders to understand (see Section 4.2 for an overview), children's most pervasive passive comprehension error involves a misinterpretation of the subject as the agent of the event denoted by the verb. Both adults and children display processing difficulty in non-canonical sentence structures (see Section 4.2.1), suggesting that the initial mapping of strings to underlying derivational structure is driven by subject-initial parsing biases. In passives, where subjects are not agents, these biases may cause especial difficulty for children. Young comprehenders are known to experience difficulty revising syntactic structures they had initially misparsed (Choi & Trueswell, 2010; Trueswell et al., 1999; Weighall, 2008). This is in part due to the order in which information appears in the sentence: child comprehenders commit to a structural interpretation of bottom-up information that is then difficult to revise due to underdeveloped cognitive control skills. Building on both the subject-initial bias and syntactic revision literature, recent work (Huang et al., 2017, 2013) suggests that children's difficulty with passives is a result of subject-initial parsing commitments and poor revision skills, which lead to the persistence of a subject-as-agent interpretation even after comprehenders encounter a cue to voice that indicates a necessity to revise.

The present work aims to investigate how children's commitments to interpreting the subject as the agent of the event described by the verb unfold over time. We exploited the flexible word order of German to investigate the impact of cue placement on subject-as-agent interpretations. One hypothesis implicit in previous

processing-based approaches to children's comprehension of passives holds that children immediately assume subjects are agents. Under this view, children extract argument roles by probabilistically aligning subjects and agents, and use highly abstract information of the form *seal-Agent* as the basis of further sentence processing. An alternative hypothesis suggests that argument role information cannot be used as the basis of further processing unless it is combined with a verb. According to this hypothesis, children still probabilistically align subjects as agents, but can only use argument role information when combined with a lexical verb into a more concrete *seal-Eater* format. These hypotheses make different predictions about when children should experience difficulty in revising a subject-as-agent interpretation. According to the subject-based hypothesis, children should experience difficulty revising a subject-as-agent interpretation anytime they encounter a subject prior to voice information, but should not experience this difficulty if voice information is provided prior to the subject. We tested this claim in Experiment 4.1, which used the V2 properties of German matrix clauses to vary the position of voice information relative to the subject. By contrast, according to the verb-based hypothesis, children should experience difficulty revising a subject-as-agent interpretation only if voice information is provided after the lexical verb. We tested this claim in Experiment 4.2, which used the verb-final properties of German embedded clauses to delay voice information until after the verb. Experiment 4.3 controlled for variation in several experimental factors that differed between Experiments 4.1 and 4.2. Overall, our results provide support for the verb-based hypothesis: German five-year-olds performed well in passive comprehension in

Experiment 4.1, regardless of whether voice information preceded or followed the subject, but showed lower passive comprehension in Experiment 4.2, where voice information was delayed until after the verb. Our findings shine a light onto the debate of whether the presence of a subject-initial bias in child comprehension entails an agent-initial bias, while showing evidence that verbs provide information that is pivotal to children's online comprehension.

4.2 Sources of variability in children's comprehension of passives

This section sketches out key sources of difficulty in children's ability to accurately parse passives. We begin with a brief outline of potential differences between children's and adults' linguistic representations, and how these might be modulated by the frequency of passives in the linguistic input. However, our main interest is difficulties with revision, and various aspects which might impact children's ability to revise an incorrect subject-as-agent misinterpretation.

Representational accounts locate the source of children's difficulties with passives in the realm of defective syntactic structure or operations. Borer & Wexler (1987) suggest that young children do not have access to A-chain formation, and that this is why children are more successful at comprehending adjectival than verbal passives. Fox & Grodzinsky (1998) observe that children's accuracy is high in comprehending *get*-passives, which do require A-chain formation, and instead suggest difficulties with theta transmission, which ensures that the verb's second theta role (agent/experiencer) is transmitted to the NP inside the *by*-phrase, as the main

source of children's passive comprehension failure. While the precise mechanics of children's deficient syntax vary, representational accounts assume that the syntactic operations required to comprehend passives become available on a particular maturational schedule, and that their ability to comprehend and produce passives prior to that point is limited.

In languages with a higher frequency of passives in the input, learners have more opportunities to learn the mapping between surface forms and derivational structure. Studies on the acquisition of Sesotho (Demuth, 1989, 1990; Demuth, Moloji, & Machobane, 2010), Inuktitut (Allen & Crago, 1996) and K'iche' Maya (Pye, 1991; Pye & Poz, 1988) note that young learners of these languages produce passives earlier than their peers in English, German and French, among others, possibly due to the higher frequency of passives in child-directed speech in these languages, or due to the better signal-to-noise ratio of mapping morphology onto passive structures (Allen & Crago, 1996). Some scholars contend that as children are susceptible to syntactic priming of passives (Messenger, Branigan, & McLean, 2012), they must have access to adult-like underlying representations of passive structures. However, in this study children produced a sizeable proportion of role-reversal errors, suggesting that even if they are able to produce surface strings corresponding to a passive construction, this may not reflect an adult-like mapping of subject to patient, and object to agent, within passives.

4.2.1 *Subject-first biases and revision*

The focus of our investigation is to explore how the order of information impacts children's subject-as-agent interpretations in passives. We first review evidence concerning the nature of children's syntactic commitments in online parsing, and their subsequent difficulty in rescinding these commitments. We then summarise findings showing that children, like adults, use a subject-first heuristic to map syntactic positions onto word order. It is harder to assess the extent to which this heuristic carries over into an interpretive commitment to the subject as an agent. Our experimental work therefore concentrates on children's differential ability to act on voice cues depending on the order of these cues relative to the subject or verb.

Both adults and children experience difficulty processing temporarily ambiguous syntactic structures, but unlike adults, children are frequently unable to revise their initial misparses once the structures are disambiguated (Trueswell et al., 1999). Trueswell et al. (1999) tested children's comprehension of relative clauses with or without overt complementisers, and found that both children and adults initially misinterpreted the prepositional phrase in reduced relative clauses (*the frog on the napkin*) as the goal of the preceding VP. For children (but not adults), this misinterpretation persisted after a disambiguating cue appeared later in the string. The misinterpretation did not occur in sentences with an overt complementiser, indicating that children's difficulty was due to processing an ambiguity, rather than being unable to parse prepositional phrases in relative clauses. Findings revealing children's lasting susceptibility to so-called garden-path effects robustly appear across languages (Choi & Trueswell, 2010) and constructions (Engelhardt, 2014). These effects may be due

to the developmental trajectory of cognitive control, which is thought to mediate conflicting stimuli including competing parses in temporarily ambiguous sentences (Mazuka, Jinocho, & Oishi, 2009; Novick, Kan, & Thompson-Schill, 2009; Novick, Trueswell, & Thompson-Schill, 2005).

Passives may be difficult to parse due to a garden-path effect caused by a mismatch of syntactic position and thematic role. Linguistic theory distinguishes between an argument's syntactic position (e.g. the subject or object of the sentence: *The dog-SUBJECT bit the man-OBJECT*) and its thematic role (e.g. the agent or theme of the verb: *The dog-AGENT bit the man-THEME*). In actives, subjects and agents coincide. In passives, on the other hand, they do not (*The man-SUBJECT-THEME was bitten by the dog-OBJECT-AGENT*). Both adults and children have a propensity to interpret initial NPs as subjects, and children's difficulty with passives may stem from the consequences of this misinterpretation. Subject-first biases have been found in many languages, but we focus on German, where additional cues in the form of case marking can prevent initial misparses in adults, but not children.

A variety of intersecting literatures provide ample evidence that adults are prone to interpreting clause-initial NPs as sentence subjects. The processing of relative clauses has proven a particularly fruitful ground for exploration. Trueswell, Tanenhaus, & Garnsey (1994) showed that adult comprehenders experience processing difficulty when encountering evidence that the temporarily ambiguous initial NP of an embedded clause is in fact its object rather than its subject. This finding replicates across a variety of contexts and languages (Bornkessel,

Schlesewsky, & Friederici, 2002; Frazier, 1987; Just & Carpenter, 1992; Konieczny et al., 1997; Ueno & Kluender, 2003).

Since German has flexible word order in which initial NPs are less often subjects than in languages with more fixed word order like English or Dutch, one might expect German comprehenders to display less pronounced subject-first parsing biases. Yet object-initial constructions are rare, making up around 3% of sentences (Bader & Häussler, 2010). Although widespread morphological syncretism in the case paradigm completely obscures the subject-object distinction in some combinations of gender and number, corpus analyses suggest that case marking nonetheless disambiguates between subjects and objects roughly 70% of the time (MacWhinney, Bates, & Kliegl, 1984). However, when case does not allow adult German comprehenders to definitively classify an initial NP as subject or object, they assume it is a subject (Bader & Meng, 1999; Knoeferle, Habets, Crocker, & Münte, 2008; Schriefers et al., 1995). Even where case information marks an initial NP as an object, electrophysiological measures show distinctive ERP profiles indicating increased processing difficulty, though this does not result in lower accuracy in behavioural measures (Bornkessel et al., 2002; Matzke et al., 2002; Mecklinger, Schriefers, Steinhauer, & Friederici, 1995). Even when case information is unambiguous and presented early, evidence from eye-tracking suggests that it may be slow to be integrated (Kamide, Altmann, et al., 2003). Finally, in ungrammatical sentences with no disambiguating case marking, German adults consistently judge the first NP to be the subject (MacWhinney et al., 1984). Although case marking provides unambiguous evidence about an NP's status as object or subject, adult

comprehenders' behaviour and electrophysiological responses nonetheless appears to favour a word order-based strategy which privileges the initial NP as a subject, as shown by persistent garden-path effects in non-canonical OS structures.

German children mirror adults in displaying a bias towards interpreting NP1 as a subject. Given unambiguous case marking, children do not display an adult-like distinctive ERP profile for initial objects before the age of 6, although even at this age their behavioural responses indicate that they interpret initial NPs as subjects (Schipke et al., 2012). This finding suggests that noticing an initial object does not entail being able to correctly interpret this information. In offline measures, children show some sensitivity to word order by age 5, although it is only at the age of 7 that they are able to prioritise case information over word order when these are in conflict (Dittmar et al., 2008). German four-year-olds' behaviour in object-relative sentence-repetition tasks is consistent with a subject-first interpretation (Diessel & Tomasello, 2005), although their success rate is higher when the head NP1 is inanimate and therefore less agent-like (Kidd, Brandt, Lieven, Tomasello, & Kidd, 2007). Children's ability to prioritise case marking over word order information develops late, and for adults, too, the use of case marking over word order is fragile in adults (Kamide, Scheepers, et al., 2003). Both adults and children are therefore susceptible to initial interpretations that identify NP1 as a subject, even in the presence of disambiguating case information.

The subject-first literature does not completely align with the processing of passives. The literature on subject-first parsing biases has tended to conflate interpreting an initial NP as a subject and interpreting it as an agent (see, for instance,

Bever, 1970). Existing studies of non-canonical object-initial structures have focused on actives, and therefore not provided empirical evidence which would allow a disambiguation between claims about subject-first or agent-first biases. Nonetheless, subject-first parsing biases provide a useful window into the steps that are involved in parsing a passive, as processing a non-canonical construction needs to undo any consequences of the comprehender's assumption that the initial NP is a subject, consequences which are plausibly severe in passives (which require non-canonical argument role assignment) just as in object-initial constructions (which require surface structure to be mapped onto a non-canonical word order). In the next section, we outline other factors that may contribute to the difficulties children may be experiencing in passive comprehension as a result of an initial subject-first parse.

4.2.2 *Factors in passive comprehension*

Correctly parsing a passive relies on three interlocking factors: the comprehender's ability to overcome the consequences of a subject-initial parse; the comprehender's ability to recognise a cue to the passive; and the comprehender's ability to act upon that cue.

Studies in English (Huang et al., 2017) and Mandarin (Huang et al., 2013) have demonstrated that reducing the strength of 5-year-olds' commitment to a subject-as-agent misinterpretation narrows the gap between their performance in active and passive comprehension. In these studies, children were presented with active or passive sentences whose initial NP was either a full NP (*the seal was quickly*

eaten by it) or a pronoun (*it was quickly eaten by the seal*). Comprehension was measured online through eye-tracking, as well as offline through an act-out task in which children's choice of toy and action determined whether they had interpreted the sentence as an active or passive. The authors hypothesise that when the subject is a pronoun, argument role assignment is delayed, resulting in a weaker agent-first commitment than when the initial subject is a full NP. Children would then interpret the pronoun according to the subject-as-agent bias, but the need to wait until after voice information was provided to identify the pronoun referent would weaken children's commitment to the subject-as-agent interpretation. Behavioural data suggest that English-speaking children do in fact comprehend passives more successfully when the initial commitment is weakened (pronoun-initial condition) than when it is not (expressed noun-initial condition; Huang et al., 2017). For Mandarin-speaking five-year-olds, the behavioural data also suggest a narrowing of the contrast in performance between active and passive comprehension, though this seems to be driven by a reduction in children's comprehension of actives, rather than an improvement in their comprehension of passives (Huang et al. 2013). Young comprehenders' performance varied depending on the available information, showing that children's comprehension of passives relies at least in part on the nature of their commitments to interpreting the subject as an agent.

Recognising a passive structure rests on the availability of cues, and their reliability as indicators of a passive. Models like the Competition Model (E. Bates & MacWhinney, 1989) or the Constraint-Based Learner model (Trueswell & Gleitman, 2004) emphasise the links between cue availability, reliability, parsing and learning:

if cues are difficult to perceive in the input, or are found in constructions other than passives, it may be difficult to take them into account in parsing, leading to comprehension errors that in turn make acquisition of the underlying representation difficult. Cue reliability and availability vary across languages. In English, passives are signalled morphologically through the use of the auxiliary *be*, the past participle of the verb, and an optional *by*-phrase. However, each of these cues appears in other active and vastly more frequent syntactic constructions. By contrast, in Mandarin, the particle *bei* is uniquely found in passives, making it a highly reliable, if rare, cue to the passive (Huang et al., 2013).

Despite considerable variation in cue reliability and availability, children's cross-linguistic difficulty in comprehending passives persists even when cue reliability is high, suggesting that other factors are at play. One possible influence is cue timing: even if a cue is available and reliable, passive comprehension may ultimately still fail if that cue appears too late to avert a commitment to an initial misinterpretation, or if a child's ability to revise is too weak to correct that misinterpretation. Cross-linguistic studies in Kannada and Tagalog (Trueswell, Kaufman, Hafri, & Lidz, 2012) show that the placement of cues to role assignment relative to other sentence information makes a key difference in children's ability to correctly interpret argument roles. In Kannada, causative morphology follows verb arguments, meaning that interpreting it correctly requires a revision of any initial subject-as-agent interpretation. In Tagalog, on the other hand, causative morphology precedes verb arguments, so a correct interpretation can guide argument role assignment. Trueswell et al. (2012) found that children learning Kannada only

infrequently interpreted intransitive sentences with causative morphology on the verb as causatives, whereas young speakers of Tagalog showed a higher proportion of interpretation as causative, suggesting that cue timing is a crucial factor in children's ability to use cues to argument role assignment.

We note that the interaction between the different factors in children's passive comprehension is obscured by considerable task-based variability in behavioural outcomes. Picture-choice tasks, on the whole, appear to elicit higher accuracy rates than act-out tasks. For instance, five-and-a-half-year-olds in Armon-Lotem et al.'s (2016) picture-choice task boasted accuracy rates of >80% in passives across eleven languages; children in the same age group performed at 66% accuracy in Huang et al.'s (2013) act-out task in Mandarin. In picture-choice tasks, distractor images frequently include a role-reversal of the target image (Armon-Lotem et al., 2016; Aschermann, Gülzow, & Wendt, 2004; Bartke, 2004). Sentence reversibility reduces children's comprehension accuracy (Turner & Rommetveit, 1967; Maratsos et al. 1985), but selecting between role-reversed images may also have focused children's attention on argument roles and boosted their overall performance. For adult speakers of German, accompanying visual information guides the interpretation of non-canonical OSV sentences in which the initial noun's morphology is ambiguous between accusative and nominative case, removing clues to thematic role assignment (Knoeferle et al. 2005), but this advantage disappears when visual information is absent (Knoeferle et al. 2008). This raises the possibility that children's comprehension may improve when visual information depicting the event being described accompanies the to-be-comprehended sentence. By contrast, act-out tasks

can reveal fine-grained detail on children's comprehension of passives due to the absence of an event depiction.

Existing studies of children's comprehension of complex syntax suggest that learners' parsing success depends in large part on their ability to access underlying representations in real time. In the present studies, we focus on how the nature of children's commitment to a subject-as-agent interpretation impacts their ability to revise their initial misinterpretations of passives. We indirectly address the issue of cue reliability by using German as our test language, in which the past tense of the auxiliary *wurde* is strongly correlated with passive structures (see section 4.3). In our investigation, we adapted a challenging act-out task (based on Huang et al. 2013, 2017) in which children were presented with possible event participants, but did not receive any visual input on possible argument role assignments within that event. Evidence from Kannada and Tagalog causative constructions suggests that cue timing of voice information relative to lexical verbs plays a role in determining children's interpretative commitments (Trueswell et al., 2012), while evidence from English and Mandarin passives highlights lexical information in the subject as a potential culprit in driving children to commit to a subject-as-agent interpretation.

Our experiments were designed to address both of these hypotheses. If children's difficulty with passive comprehension is due to not being able to revise a subject-as-agent commitment made at the subject, children should comprehend passives well when the voice cue precedes the subject, but poorly if the voice cue follows the subject. By contrast, if children's difficulty with passive comprehension is due to the combination of argument role information with verb information, one

would expect a pattern of results in which children's performance in passive comprehension drops when the voice cue is presented after the lexical verb. Experiment 4.1 explored the subject-based hypothesis by using a sentence-initial adverb to trigger the V2 properties of German matrix clauses, such that a highly reliable cue to the passive could be manipulated to appear either before or after an expressed subject. Experiment 4.2 explored the verb-based hypothesis by using the verb-final properties of German subordinate clauses, placing a highly reliable cue to the passive after the main verb. Experiment 4.2 departed significantly from Experiment 4.1 in that it tested children on short, not long, passives. Experiment 4.3 therefore controlled for the possibility that the absence of pronouns, rather than the placement of the voice cue relative to the main verb, was responsible for the experimental outcomes of Experiment 4.2.

4.3 German passives: a preliminary corpus study

As in many other languages, passives in German are exceedingly rare, making up >0.5% of recorded utterances in the dense Manchester-Leipzig corpus (Abbot-Smith & Behrens, 2006). This is comparable to the vanishingly low rates of passives in English (Roland, Dick, & Elman, 2007) and Mandarin (Huang et al., 2013). While the flexible word order of German permits object-initial sentences, these constructions are highly infrequent. Object-initial clauses make up only around 3% of all sentences in written corpora (Bader & Häussler, 2010), though in child production they make up 15% of all V2 utterances (Poeppl & Wexler, 1993). Adverb-initial sentences are

robustly represented in child speech, making up roughly 10% of one two-year-old's V2 production (Poeppel & Wexler, 1993). As in many other languages, the acquisition of passives in German is delayed, with production becoming more robust after the age of four (Mills, 1985). Importantly, the auxiliary *werden* is used both to introduce passives and the future tense. Adult comprehenders rely on temporal information from adverbs to disambiguate these two uses of *werden* (Knoeferle et al., 2005). Despite competition between these two uses of *werden*, with vastly higher input frequency for the future than for the passive use, Abbot-Smith & Behrens's (2006) study of the Manchester-Leipzig corpus revealed that Leo first started producing *werden*-passives around the age of two, productively using it with around 60 different verb types by the age of three.

We conducted a follow-up corpus study to investigate the frequency of passives with the auxiliary *werden* in caretaker speech using the dense Manchester-Leipzig Leo corpus (Behrens, 2006). The results, displayed in **Error! Unknown switch argument.**, show considerable variability in the degree to which a specific conjugation of *werden* correlates with utterances in the passive. Of all instances of *werden* in the corpus, less than one-third are passives. However, depending on person and number features, instances of *werden* vary widely in terms of their likelihood of appearing in a passive. Of total passive occurrences, over 90% are divided between just three forms of *werden* in the present tense, but less than half of the instances of these forms correspond to passives overall. Although past tense passives are less frequent overall, past tense conjugations of *werden* have an 80-100% chance of appearing in a passive. This makes *wurde* (3rd pers. sg. past) a highly reliable cue to

passive, and we therefore used this form in our target stimuli (see Materials sections for details). Existing studies of German (Armon-Lotem et al., 2016; Aschermann et al., 2004; Bartke, 2004) use the present tense of *werden*. Children may therefore encounter the cue to voice early in the sentence, but due to its ambiguity, do not have reliable evidence of the passive until they encounter the past participle. This ambiguity may contribute to variability in existing studies of children’s performance in passive comprehension.

Table 4.1: Occurrences of *werden* in caretaker speech

person	tense	form	occurrences (absolute)	passives (absolute)	% passive	% of total passives
1 st SG	present	werde	230	3	1.3	0.3
2 nd SG	present	wirst	164	3	1.8	0.3
3 rd PL	present	wird	1,902	447	23	47
1 st , 3 rd PL	present	werden	1,015	445	44	47
2 nd PL	present	werdet	10	0	0	0
1 st , 3 rd SG	past	wurde	49	42	86	4
2 nd PL	past	wurdest	4	4	100	0.4
1 st , 2 nd , 3 rd PL	past	wurden	10	8	80	.8
Total			3,384	952	28	100

In a second corpus study, we investigated the distribution of the past participles used in the three experiments in 42-billion word corpus of written German (Das Deutsche Referenzkorpus DeReKo) in order to determine the balance between active and passive use of these verbs, to avoid any lexical frequency bias that might affect

processing. Even more than in English, the formal register of written German has a higher passive frequency than the spoken language. Of the over 3 million hits for one or more of the selected past participles, 39% co-occurred with *werden* and 45% co-occurred with *haben* (the active auxiliary which is used to form the past perfect), indicating that these verbs were not biased towards active or passive usage in any way. Hand-coding of the first 100 instances in each case revealed that 85% of instances of the selected past participles co-occurring with *werden* were passives in this corpus, while 86% of *haben* co-occurrence instances were actives, indicating that the rough metric of co-occurrence with *haben/werden* is a reliable gauge of whether a verb is likely to be used in the active or passive voice. Overall, these corpus data show that the past participles used in these studies were roughly equally likely to appear in active and passive sentences. Given the unusually high incidence of passives in written German, this is likely an overestimation of the frequency with which children might encounter these verbs in the passive in natural child-directed speech. However, these numbers indicate that these verbs at best appear equally in actives and passives, suggesting that there should be no frequency bias towards either voice in processing these verbs.

4.4 Experiment 4.1: Subject/Voice Manipulation

Experiment 4.1 chiefly investigated the importance of cue placement in children's ability to revise subject-as-agent interpretations and therefore succeed in passive comprehension. The main hypothesis under investigation claims that children's

ability to revise a passive relies crucially on the placement of cues to the passive. In German, where the only obligatory cue to passive is the auxiliary, V2 syntax allows for a tightly controlled manipulation of cue placement. If providing a cue to voice prior to the subject allows the parser to avoid a subject-as-agent misinterpretation that is difficult to revise, children should be more successful at comprehending passives than when the cue to voice occurs after the subject.

Following Huang et al. (2013, 2017), Experiment 4.1 varied the position of the cue to voice with respect to the subject in a 2x2 design, in which VOICE was manipulated within and CUE TIMING between subjects. We tested CUE TIMING between, rather than within, subjects because pilot testing for Huang et al. (2013) suggested that within-subject variation in CUE TIMING led to a greater number of trials lost due to participants picking up all toys, not completing the action, etc., suggesting that higher variability in test stimuli may have led to confusion. Participants' task was to select toys from a display and use them to act out aurally presented stimuli. Video cameras were used to capture eye movement data and to record participants' actions.

4.4.1 Participants

48 5-year-olds (mean age 5;0, range: 3;10-6;8, SD: 0;10; 28 boys) were recruited from the U Potsdam BabyLab subject pool, or from a daycare centre in central Berlin. All children received a small reward for their participation; families of children recruited through U Potsdam also received €7.50 in compensation for travel expenses to the test location. There were no significant age differences between participants tested in different CUE TIMING conditions ($t = .845, p=0.4$). Two further child

participants opted out after the start of the experiment, and their data were not considered for any further analysis. IRB approval for these studies was granted by the University of Maryland, College Park, the University of Potsdam, and the ethics commission of the Deutsche Gesellschaft für Sprachwissenschaft.

4.4.2 *Materials*

Twelve test stimuli were constructed around sets of three toys: the subject (e.g. a seal), a likely theme (e.g. a fish) and a likely agent (e.g. a shark). Test sentences contained an expressed noun as the subject, and a pronoun as an object (direct object in actives; within a by-phrase in passives). All lexical verbs in test stimuli in Experiment 4.1 and subsequent experiments were actional verbs. Children’s assignment of argument roles to the subject, and therefore their ability to identify and use cues to voice, was indexed by their choice of toy to depict the pronoun, as well as the action itself.

In German, the cue to passive is a finite auxiliary verb (*hat*, active vs. *wurde*, passive) and appears in V2 position in matrix clauses. In the Subject-Before-Voice condition, the voice head occurred after the expressed noun. In the Voice-Before-Subject condition, the voice head preceded both noun phrases due to the insertion of a temporal adverb in sentence-initial position. (See Table 4.2 for a sample stimulus set.)

Table 4.2: Experiment 4.1 sample stimuli

Condition	German sentence	Translation
pre-subject active	Heute hat die Robbe ihn gefressen. today has the seal it.acc eaten	‘The seal ate it today.’

pre-subject passive	Heute wurde die Robbe von ihm gefressen. today was the seal by it.dat eaten	‘The seal was eaten by it today.’
post-subject active	Die Robbe hat ihn heute gefressen. the seal has it.acc today eaten	‘The seal ate it today.’
post-subject passive	Die Robbe wurde heute von ihm gefressen. the seal was today by it.dat eaten	‘The seal was eaten by it today.’

In all test stimuli, the likely agent and likely theme agreed in grammatical gender, so as to ensure that participants’ choice was not based on gender-matching between pronouns and referents.

Since the ultimate choice of pronoun referent relied on world knowledge as well as interpretation of the voice cue, two norming studies were conducted via Amazon Mechanical Turk. The likelihood norming study tested the agentive relationship of the three toys to each other. Participants (n=60) were presented with two items from each stimulus set (either expressed noun and likely agent, or expressed noun and likely theme) and asked to judge how likely one object was to do something to the other. Participants judged expressed nouns (e.g. seal) as more likely to do something to likely themes (fish; mean: SD: 1.06) than likely agents (shark; mean: SD: 1.36; $t=33.072$, $p<.001$), and likely agents (shark) as more likely to do something to expressed nouns (seal; mean rating: 4.57, SD: 2.21) than likely themes (fish; mean rating: 3.53, SD: 1.87; $t=4.83$, $p<.001$). The relatedness norming study tested whether associations between items could be a factor in driving participants’ choice of toys. Participants (n=60) rated the relatedness of two out of three objects in each stimulus set (expressed noun and either likely agent or likely theme) on a scale of 1 (“not at all related”) to 7 (“extremely related”). Judgements of the relatedness of

expressed nouns to likely agents (mean: 4.09, SD: 1.96) differed significantly from judgements of the relatedness of expressed nouns to likely themes (mean: 3.29, SD: 1.8, $t=5.74$, $p<.001$). However, relatedness did not impact statistical outcomes for act-out or eye-movement data.

To ensure that eye-gazes were not driven by the toys' relative sizes, the named item (e.g. the seal) was always smaller than the likely agent (e.g. the shark) and larger than the likely theme (e.g. the fish). The location of the toys was rotated between test trials to ensure that looks towards certain objects were not due to biased looks towards particular display regions.

Test items were randomly distributed across four experimental lists in a Latin square design. The twelve test sentences were mixed with 36 filler trials. Of these, twelve used the same toy sets as the test sentences and were presented immediately after a test stimulus, and twenty-four further sentences were used with twelve new sets of three toys, which were interleaved with test stimulus sets. Fillers referred directly to one or two participants, using either conjoined subjects (*Der Junge und der Vater haben sich gedreht*, “the boy and the father spun around”), ambiguous reflexives that could be interpreted as either transitive or intransitive (*Die Prinzessin und der Frosch haben sich gewaschen*, “the princess and the frog washed themselves/each other”), or singular subjects with transitive verbs (*Der Frosch ist auf und abgehüpft*, “the frog jumped up and down”). As in test sentences, filler verbs were in the present perfect tense, such that auxiliaries were presented in V2 position and the lexical verb appeared sentence-finally.

All sentences were recorded by a female native speaker of German and enunciated clearly in child-directed speech.

4.4.3 *Procedure*

Participants sat or stood within arm's reach of a visual display. Prior to each trial, the experimenter (a native speaker of German) placed each toy individually on the display and labelled it ("This is a seal. This is a shark..."). Participants were instructed to look straight ahead at the display and act out the story they had heard using the toys on the display. Participants did four practice trials prior to the start of the experiment, which they were allowed to repeat until they were comfortable with the task. During the task, participants were filmed using two camcorders: one positioned within the display monitored participants' eye movements, another positioned next to participants recorded their actions.

4.4.4 *Results*

Behavioural data were analysed according to whether the child had assigned the correct argument role to the subject (agent in actives; theme in passives).² This was

² This is a departure from the coding scheme laid out in Huang et al. (2013, 2017). According to those authors' scheme, if participant's action correctly identified the two participants and their relation to each other (e.g. making the seal eat the fish, or the shark eat the seal), this was coded as "correct"; correctly identifying the two participants, but incorrectly portraying their relationship (e.g. making the seal eat the shark) was coded as a "role reversal"; and portraying some other action was coded as "incorrect." We initially coded data from Experiment 4.1 according to this scheme, and the statistical analyses and generalisations do not differ from the generalisations we draw here. However, we adopt the simpler coding scheme outlined above to provide continuity with Experiments 4.2 and 4.3, whose

identified through the direction of their actions (e.g. was the seal moving towards the shark or vice versa?), the exclamations they ascribed to different toys (“Yum! What a tasty seal!” while moving the shark), explanations while moving toys (“Seals really like to eat fish, you know”), and, to a lesser extent, their choice of second toy (did the seal eat the fish or the shark?). As some participants did “silly” act-outs in which agents did things to creatively chosen themes, but the role assignment to the subject was abundantly clear, we chose to code these actions as correct. Overall, actions were counted as “correct” if the subject was assigned the correct role (agent in actives, theme in passives), “reversed” if the subject was assigned the reversed role (theme in actives, agent in passives), or “other error” for any other type of error.

One participant was excluded from further analysis due to incorrect actions in over 50% of active trials. <1% of trials were excluded from further analysis due to experimenter or technical error. Data from one stimulus item were removed from further analysis because over 50% of responses in the active conditions were incorrect across participants.

Eye-tracking data were analysed by trained research assistants using a frame-by-frame viewing software, Vcode (Hagedorn, Hailpern, & Karahalios, 2008), to code participants’ looks from stimulus onset to the start of their action. Research assistants did not speak German and were blind to trial condition and object location. Blinks, looks away from the display, and instances when the child’s eyes were obscured were coded as track loss and excluded from further analysis, accounting for

test sentences do not include objects and where the only measure of accuracy is role assignment to the subject.

10% of the data. One participant was excluded from eye-tracking analysis due to data loss from equipment failure, and one due to poor lighting.

All statistical analyses were conducted using the open-source software R (R Core Team, 2017) and the analysis packages lme4 (D. Bates, Maechler, Bolker, & Walker, 2015), effsize (Torchiano, 2017) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017).

4.4.4.1 Behavioural data

Error! Unknown switch argument. provides a summary of average rates for children's response types, by condition. Children's performance in all four conditions was significantly higher than chance, estimated conservatively at 50% (one-sample t-tests; all p-values <.001). For further statistical analysis, actions were coded as a binary value based on accuracy (1 for a correct action, 0 for an incorrect action). As stated in the Materials section, a second analysis of relatedness data from the norming study revealed a significant effect of relatedness between subject and likely theme vs. subject and likely agent. To assess the extent to which relatedness impacted children's actions, we computed the relatedness difference between ratings of the subject's relatedness to the likely theme (i.e. seal vs. fish) and ratings of the subject's relatedness to the likely agent (i.e. seal vs. shark) for each item. Negative relatedness difference values indicate a stronger bias towards considering the agent related, whereas positive values indicate a stronger bias towards considering the theme related to the subject.

Table 4.3: Behavioural results, Exp. 4.1

Condition	% Correct (SE)	% Reversed (SE)	% Other Incorrect (SE)
Pre-Subject Active	87 (3)	9 (2.5)	3 (1.5)
Pre-Subject Passive	86 (3)	10 (2.5)	3 (1.5)
Post-Subject Active	88 (3)	9 (3)	2.5 (1.5)
Post-Subject Passive	82 (3.5)	14 (3)	3.5 (1.5)

A logistic mixed-effects model with voice, cue timing and relatedness difference as fixed effects, and subjects and items as random effects, revealed that accuracy was not significantly affected by fixed effects VOICE ($z = .18, p > .5$), CUE TIMING ($z = -.2, p > .5$), or RELATEDNESS DIFFERENCE ($z = .2, p > .5$). RELATEDNESS DIFFERENCE was therefore not included as a factor in any further statistical analyses. Hedges's g was calculated over children's average accuracy by voice, yielding an effect size of 0.17.

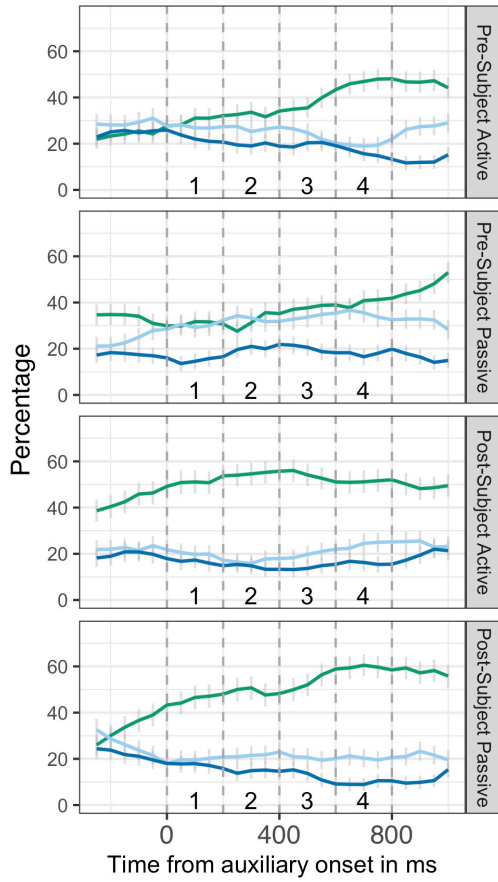
4.4.4.2 Eye-tracking data

Our behavioural analysis rested on whether children had assigned the correct argument role to the subject (depending on whether the seal was portrayed as an agent or a theme). We found that children varied in terms of whether they selected the likely theme or agent as their second event participant. We therefore evaluated children's gazes according to their actions: for instance, if a child had correctly acted out a passive sentence in which the seal got eaten by the fish, we analysed looks towards the fish as looks towards the target, whereas in other trials in which a child correctly acted out a passive using the shark to eat the seal, looks towards the shark

were evaluated as target looks. This allowed us to calibrate our eye movement analysis to take children's actual choice of target and distractor into account.

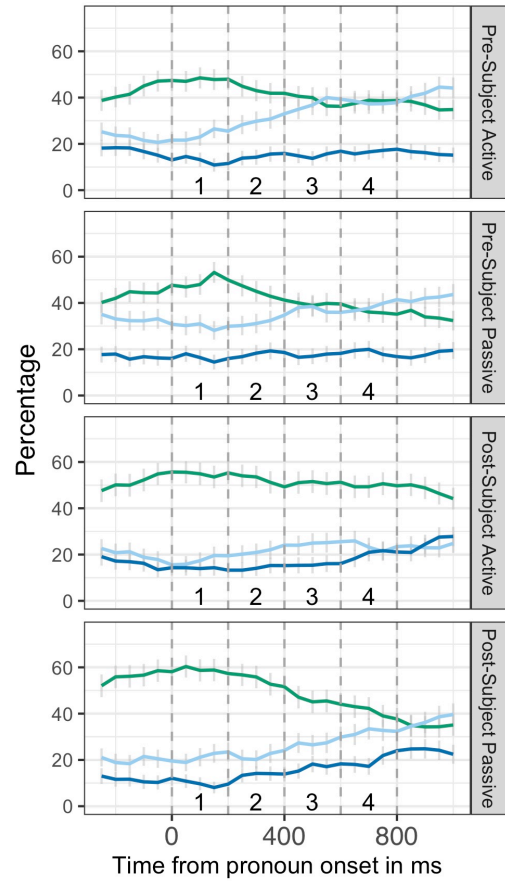
The earliest possible point at which comprehenders' gazes towards targets (likely themes in actives i.e. fish-type objects, likely agents in passives i.e. shark-type objects) or distractors (vice versa) could reflect role assignment is the onset of the voice head. **Error! Unknown switch argument.** shows the time-course of gazes towards targets, distractors and subjects (seal-type objects; calculated over total looks including looks to the centre or the empty box) time-locked to the voice head onset, with no shifting to account for saccade planning. Eye movements were time-locked exactly to the timing of different words in the stimulus. Previous literature estimates a saccade-planning buffer of 200 ms for adults (Matin, Shao, & Boff, 1993) or 400 ms for children (Huang et al., 2013, 2017). However, these estimates assume that different types of linguistic information are processed at uniform speeds, which seems implausible given their varying complexity. By time-locking eye movements exactly to the timing of the input in our graphs, we provide a more conservative visual measure of comprehenders' responses to cues.

A second region of interest in the eye movement analysis was the onset of the pronoun or by-phrase, which was a second disambiguation point within the sentence. Figure 4.4 shows the time-course of looks towards targets, distractors and subjects from the onset of the pronoun/by-phrase. Table 4.4 shows the average duration of each of these regions of interest. (In both auxiliary and pronoun regions, the active condition is monosyllabic, whereas the passive is bisyllabic.)



Looks: — Subject — Target — Distractor

Figure 4.1: Eye movements after auxiliary onset (Exp. 4.1)



Looks: — Subject — Target — Distractor

Figure 4.2: Eye movements after pronoun onset (Exp. 4.1)

Table 4.4: Average duration of regions of interest, Exp. 4.1

Condition	Auxiliary	Pronoun
Pre-Subject Active	234 ms	207 ms
Pre-Subject Passive	351 ms	435 ms
Post-Subject Active	232 ms	181 ms
Post-Subject Passive	349 ms	424 ms

We assessed participants' eye movements in four bins of 200 ms following the onset of each region of interest. Dividing each region into shorter bins provides a more fine-grained view of the time-course of gazes across each window. The analysis was done in four bins because, depending on condition, the two regions of interest varied between roughly 200 and 400 ms in duration. In addition, since comprehenders require some time to integrate linguistic information and plan saccades accordingly, we included a further two time bins in the analysis.

Fixation preferences were calculated across 200-ms bins by items and subjects in order to assess the extent to which children's target and distractor fixations differed as a function of VOICE and TIMING. In passives, average looks to distractors were subtracted from average looks to targets, such that more positive values indicated a preference for the likely agent (and therefore a sensitivity to the passive cue). In actives, average looks to targets were subtracted from average looks to targets, such that more negative values indicated a preference for the likely agent (and therefore sensitivity to active cues). Trials in which there were no looks to targets or distractors, or where looks to targets and distractors were equal, were retained in the analysis, as this otherwise would have resulted in a 40% loss of data points. A separate linear

mixed-effects model listing participants and items as random factors, and VOICE (active/passive) and TIMING (pre-subject/post-subject) as fixed factors, was run for each 200-ms time bin from the Auxiliary onset; results are reported in **Error! Unknown switch argument.** To correct for the possibility of type II error, a Bonferroni correction was applied (using a factor of 4, as there were four bins analysed in this region); adjusted *p*-values (i.e. multiplied by four) are provided in the table.

Table 4.5: Eye-movement statistics after auxiliary onset, Exp. 4.1

	Bin 1	Bin 2	Bin 3	Bin 4
	(0-200 ms)	(200-400 ms)	(400-600 ms)	(600-800 ms)
<i>t</i> value, Voice	0.6	0.86	1.2	2.2
adjusted <i>p</i> -value, Voice	> 0.5 n. s.	> 0.5 n. s.	> 0.5 n. s.	> 0.1 n. s.
<i>t</i> value, Timing	- 0.2	-0.57	0.09	0.66
adjusted <i>p</i> -value, Timing	> 0.5 n. s.	> 0.5 n. s.	> 0.5 n. s.	> 0.5 n. s.
<i>t</i> value, interaction, Timing x Voice	1.15	1.18	0.76	0.23
<i>t</i> value, interaction, Timing x Voice	> 0.5 n. s.	> 0.5 n. s.	> 0.5 n. s.	> 0.5 n. s.

The same analysis was adopted for four 200-ms time bins starting from the onset of the Pronoun region. Results are reported in **Error! Unknown switch argument.**, with a Bonferroni correction for a factor of four as in the Auxiliary analyses.

Table 4.6: Eye-movement statistics from pronoun onset, Exp. 4.1

	Bin 1	Bin 2	Bin 3	Bin 4
	(0-200 ms)	(200-400 ms)	(400-600 ms)	(600-800 ms)
<i>t</i> value, Voice	-0.4	2	2.5	2.2
adjusted <i>p</i> -value, Voice	> .5 n. s.	> .1 n. s.	0.045 *	0.12 n. s.
<i>t</i> value, Timing	-1	-1.1	-1.2	-1.7
adjusted <i>p</i> -value, Timing	> .5 n. s.	> .5 n. s.	> .5 n. s.	> .5 n. s.
<i>t</i> value, interaction, Timing x Voice	1.1	1.1	1.5	1.7
<i>t</i> value, interaction, Timing x Voice	> .5 n. s.	> .1 n. s.	> .5 n. s.	> .5 n. s.

4.4.5 Discussion

Across online and offline measures, Experiment 4.1 revealed that German-speaking five-year-olds are remarkably adept at correctly assigning argument roles in active and passive sentences. The experimental hypothesis we pursued in this experiment stated that if children's difficulty in passive comprehension stems from an initial subject-as-agent misinterpretation occurring at the subject, this misinterpretation should be avoidable if cues to voice are provided prior to the subject. However, act-out results showed that participants were able to avoid the effects of any misinterpretation even when the cue to voice was provided after the initial subject. Eye movement results across two regions of analysis showed no consistent preference for looks to the target (shark-type objects in passive, fish-type objects in active conditions) over looks to the distractor (fish-type objects in passive, shark-type objects in active conditions), with the exception of one analysis bin during the

pronoun region where there was a statistically significant contrast in fixation preference. We discuss the impact of eye movement data further in Section 4.7, but note for the time being that the eye movement results of Experiment 4.1 do not show conclusive evidence of differential argument role interpretation between active and passive sentences.

In Experiment 4.2, which used the same methodology as Experiment 4.1, we explored an alternative hypothesis which holds that young German comprehenders' ability to correctly interpret passives depends on the nature of the argument role information they have received. In particular, this hypothesis suggests that children's commitment to a subject-as-agent interpretation becomes difficult to revise once argument role information is combined with verb information in a *seal-eater* format. Experiment 4.2 used the verb-final properties of German embedded clauses to isolate the contribution of the lexical verb to online argument role assignment in children's passive comprehension. Since the main cues to voice in these stimuli, auxiliaries *hat* and *wurde*, are finite verbs, they follow the lexical verb in embedded clauses. If the main difficulty in passive comprehension is a commitment to subject-as-agent role assignment occurring at the lexical verb, rather than a subject-as-agent misinterpretation occurring at the subject, German children should experience considerably more difficulty in interpreting passives in which voice cues follow the lexical verb, as in other languages, than they did when voice cues preceded the lexical verb. In addition, we tested the extent of German children's subject-as-agent parsing biases at the subject by using a conjoined verb phrase. If German learners are biased

towards interpreting a subject as an agent, looks to the likely theme or the subject should outweigh looks to the likely agent.

4.5 Experiment 4.2: Verb/Voice manipulation

4.5.1 *Materials*

Test stimuli in Experiment 4.2 were constructed around the same object sets as in Experiment 4.1. Sentences included the matrix phrase “Zeig’ mal, wie...” (“Show me how...”) and an embedded clause describing an event, with an expressed subject (e.g. the seal) but no object (see **Error! Unknown switch argument.** for examples). The requirements of German syntax would have forced the object or by-phrase to appear prior to the verb, undermining our experimental paradigm. Without the object or by-phrase, children were forced to rely on auxiliaries *hat* and *wurde* for voice information, and received this information only after the lexical verb. Test stimuli in Experiment 4.2 were therefore short passives, a departure from the long passives used in Experiment 4.1. Despite the oft-cited finding that children fare better in comprehending short passives than long passives, this result is highly variable: it appears only in non-actional passives (Fox & Grodzinsky, 1998), does not extend into a statistically robust long-short passive performance contrast despite high numeric contrasts in performance (Gordon & Chafetz, 1990; Hirsch & Wexler, 2006; Maratsos et al., 1985), appears only under application of unusual and inappropriate statistical tests on small sample sizes (Harris, 1976) or only in a small subset of

participants, with a low number of trials per participant per condition (Fox & Grodzinsky, 1998). If the results of Experiment 4.2 differ from those of Experiment 4.1 due to the use of short passives, we would expect, if anything, an improvement in children’s comprehension of passives.

Table 4.7: Experiment 4.2 sample stimuli

Condition	German Sentence	Translation
post-verb active	Zeig’ mal [wie die Robbe gebissen und gefressen <u>hat</u>] show PRT [how the seal bitten and eaten has]	‘Show (me) how the seal bit and ate.’
post-verb passive	Zeig’ mal [wie die Robbe gebissen und gefressen <u>wurde</u>] show PRT [how the seal bitten and eaten was]	‘Show (me) how the seal was bitten and eaten.’

Verb regions consisted of conjoined verbs with related meanings (e.g. *gebissen und gefressen*, “bitten and eaten”). They were optionally transitive, so as to yield an acceptable sentence in active conditions even in the absence of a second argument. Experiment fillers were adapted from Experiment 4.1 to match the syntactic structure of test stimuli. Experimental lists and audio stimuli were prepared as in Experiment 4.1. The single experimental parameter, VOICE, was varied within subjects as in Experiment 4.1. In order to reduce the likelihood of experimenter error, two experimental lists were compiled such that active and passive test sentences alternated (separated by three filler items as in Experiment 4.1), but the order of test items was identical in both experimental lists.

4.5.2 *Participants*

24 German-speaking 5-year-olds (mean age: 5;0, SD: 0;6; 14 boys) were recruited from daycare centres in Berlin. No participant had taken part in Experiment 4.1. Data from two further participants were collected but later excluded due to multilingualism. An additional three participants' data were partially collected but discarded due to technical disruptions or early experiment termination. There were no differences in participants' ages across Experiment 4.1 and 4.2 ($p > .5$).

4.5.3 *Procedure*

Experimental procedure was identical to Experiment 4.1 procedure.

4.5.4 *Results*

4.5.4.1 Behavioural data

We replicated the analysis from Experiment 4.1. One test item was excluded from further analysis due to participants' poor performance in active conditions. <1% of test trials were excluded from analysis due to experimenter or technical error. Some children performed only one of the actions described in the conjoined VP, but others performed two in sequence. Where this was the case, we evaluated the argument role assignment they demonstrated in the first action.

Table 4.8 shows participants' performance in Experiment 4.2. In both active and passive conditions, participants scored at above-chance accuracy levels (chance again estimated conservatively at 50%; all p values $< .001^{***}$).

Table 4.8: Behavioural results, Exp. 4.2

Condition	% Correct (SE)	% Reversed (SE)	% Other Incorrect (SE)
Post-Verb Active	94 (2)	3 (1.5)	3 (1.5)
Post-Verb Passive	66 (4)	29 (4)	5 (2)

A generalised mixed-effects model with participants and items listed as random intercepts showed a performance contrast by VOICE ($z = -5.1, p < .001^{***}$), with significantly lower performance in passive conditions. A further generalised mixed-effects model, also with participants and items listed as random intercepts, confirmed that participants' incidence of committing a role-reversal error increased in passives ($z = 4.8, p < .001^{***}$). A final generalised mixed-effects model (random intercepts: participants and items) including fixed factors VOICE and AGE (in months) yielded no statistical effect of participants' age on their act-out accuracy ($p > .3, z = .99$). Hedges's g was calculated over children's average accuracy by voice, yielding an effect size of 0.83.

4.5.4.2 Eye-tracking data

Analysis of eye movements was the same as in Experiment 4.1. The two regions of interest were the conjoined verb region and the auxiliary, which was the final word in

the sentence (see **Error! Unknown switch argument.** for average durations by condition). Figure 4.3 shows participants' looks towards subjects, targets and distractors after the onset of the verb region, and Figure 4.4 shows the same measures after the onset of the auxiliary.

Table 4.9: Average duration of regions of interest in Experiment 4.2

Condition	Conjoined Verb	Auxiliary
Post-Verb Active	1477 ms	560 ms
Post-Verb Passive	1456 ms	653 ms

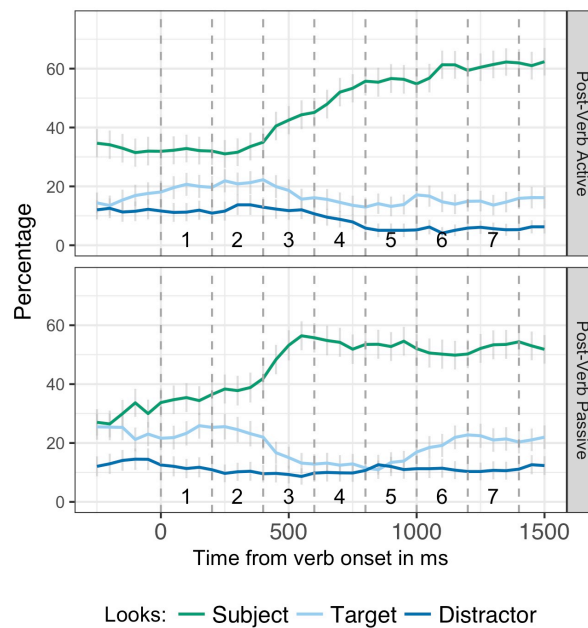


Figure 4.3: Eye movements after verb onset (Exp. 4.2)

Table 4.10: Eye-movement analysis from verb onset, Exp. 4.2

Time Window	<i>t</i> value	<i>p</i>-value (Bonferroni correction: 7)
1 (0-200 ms)	2.1	> .2 n.s.
2 (200-400 ms)	2.3	> .1 n.s.
3 (400-600 ms)	1.9	> .3 n. s.
4 (600-800 ms)	1.4	> .5 n.s.
5 (800-1000 ms)	1.3	> .5 n.s.
6 (1000-1200 ms)	2.9	.03 *
7 (1200-1400 ms)	3.2	.012**

As before, regions of interest were statistically analysed in 200 ms time windows. As the average duration of the verb region was around 1400 ms, eye gazes were evaluated in seven consecutive bins. For the auxiliary region, eye gazes were evaluated in five bins: three that covered the auxiliary itself, and two that extended past the end of the stimulus. Statistical analyses again followed the eye-movement analysis used for Experiment 4.1. For each region, and for each time window, a separate linear mixed-effects model was applied to fixation preference scores, listing participants and items as random factors, and VOICE (active/passive) and TIMING (pre-subject/post-subject) as fixed factors. Results for the verb region are reported in Table 4.10 (*p*-values Bonferroni corrected by a factor of 7) and for the auxiliary region in Table 4.11 (*p*-values Bonferroni corrected by a factor of 5).

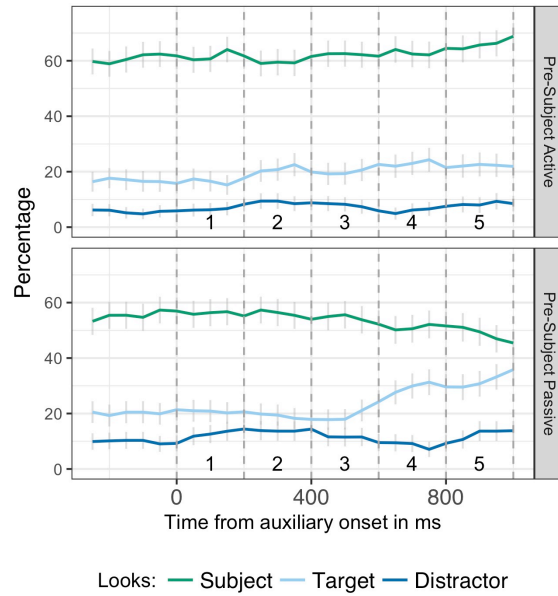


Figure 4.4: Eye movements from auxiliary onset (Exp. 4.2)

These results show that children’s fixation preferences differed towards the end of the verb region, carrying over into the first bin of the auxiliary region. In the auxiliary region itself, the first and last analysis bins show a significant contrast in fixation preferences by voice. These results are puzzling. Assuming that at least 200 ms are required for comprehenders to plan a saccade (Matin, Shao, & Boff, 1993), and that this is further delayed for child comprehenders (Huang et al., 2017), these results seem to indicate that children encountered cues to begin generating differential eye movements by the middle of the verb region (around 600-800 ms after the VP onset), then abandoned this preference towards the end of the verb region. The differentiation of eye movements by voice is expected for the final analysis bin in the auxiliary region, since this would coincide with eye movement generation towards the middle or end of the auxiliary itself.

Table 4.11: Eye movement statistics from auxiliary onset, Exp. 4.2

	Bin 1	Bin 2	Bin 3	Bin 4
	(0-200 ms)	(200-400 ms)	(400-600 ms)	(600-800 ms)
<i>t</i> value, Voice	2.7	2.1	1.9	4
adjusted <i>p</i> -value, Voice	.35 *	> .1 n.s.	> .2 n.s.	> 0.001***

4.5.5 Discussion

Experiment 4.2 tested the distinction between subject-as-agent commitments made at the subject and those made at the verb by withholding voice information until after the lexical verb. Building on the results of Experiment 4.1, we hypothesised that subject-as-agent commitments made at the subject in the form *seal-agent* might be easy to revise (hence children’s overall success in passive comprehension in Experiment 4.1, regardless whether voice information preceded or followed the subject), whereas such commitments might be harder to revise if made at the lexical verb in the form *seal-eater*. In Experiment 4.2, we found that children’s accuracy in assigning an argument role to the subject dropped steeply; this was because participants now committed significantly more role reversal errors, especially due to incorrectly interpreting the subject of a passive sentence as an agent and failing to revise this error. Children’s eye movements revealed significant contrasts in fixation preferences during the verb region (discussed further in section 4.7), and within 400 ms of hearing voice information, children had oriented towards the correct target.

This contrasts with behavioural outcomes, which show a persistent subject-as-agent interpretation, regardless of sentence voice.

Our findings indicate that the argument role commitments children make at the lexical verb are difficult to revise, even if they generate correct eye-movements and even if late-arriving cues lead participants to direct gazes towards the appropriate target based on voice. Children's poor passive comprehension in Experiment 4.2 stands in stark contrast with their high performance in 4.1, whose participant population was comparable in every respect. However, several contrasting choices in experimental design make it difficult to directly compare the results of Experiments 4.1 and 4.2. Experiment 4.2 used conjoined verb phrases, rather than simple verb phrases as in Experiment 4.1. Children's undiminished facility in correctly acting out active test sentences and, anecdotally, fillers suggests that conjoined verb phrases did not place undue strain on their comprehension. It is conceivable, however, that parsing a conjoined verb phrase placed a processing burden on child comprehenders that resulted in a higher incidence of comprehension errors in passives, which may require greater resources in processing than actives due to the need for revision. In addition, Experiment 4.2 used short passives, unlike the long passives in Experiment 4.1. As discussed in section 4.5.1, the evidence for a difference in children's comprehension outcomes between long and short passives is variable, and, if anything, would have improved children's comprehension overall. Experiment 4.3 controlled for these two possibilities by replicating the Subject-Voice conditions of Experiment 4.1 with conjoined verb phrases, but no objects. We hypothesised that if either or both of these factors impact children's performance in passive

comprehension, participants in Experiment 4.3 would perform markedly worse in passive than in active sentences. By contrast, if children's performance in Experiment 4.3 reveals no differences in comprehension accuracy for actives over passives, the differences in children's passive comprehension between Experiments 4.1 and 4.2 can more confidently be attributed to differences in the ordering of lexical verb and voice information and therefore, point towards qualitative contrasts in argument role commitments depending on the availability or absence of information from the lexical verb as the underlying source of children's difficulty with passives.

4.6 Experiment 4.3: Control

4.6.1 *Materials and Procedure*

Materials were again constructed using the same object sets as in Experiments 4.1 and 4.2. Test sentences were matrix clauses with the cue to voice (auxiliary *hat* or *wurde*) in second position, followed by a conjoined verb phrase (see Table 4.12 for examples). Sentences did not contain an object, i.e. they were either active transitive sentences or short intransitive passives. As in Experiment 4.2, fillers were adapted from Experiment 4.1, although due to concerns that some children were uncomfortable performing some of the actions expressed in these sentences (hitting, kicking, etc.), five out of the 36 fillers were altered to depict less violent events. Four experimental lists were constructed to vary the parameter VOICE within participants, with the same order of test items across two lists and a reversed order for test items

across two other lists (to allow for later analysis of first-half vs. second-half performance differences). The procedure for recording stimuli and conducting experiments were identical to the procedures applied in Experiments 4.1 and 4.2.

Table 4.12: Experiment 4.3 sample stimuli

Condition	German Sentence	Translation
post-subject active	Die Robbe <u>hat</u> heute gefressen. the seal has today eaten	‘The seal ate (it) today.’
post-subject passive (revision required)	Die Robbe <u>wurde</u> heute gefressen. the seal was today eaten	‘The seal was eaten (by it) today.’

4.6.2 *Participants*

24 monolingual German children (mean age: 5;02, range: 4;8 – 6;6, SD: 0;6; 9 boys) were recruited from the U Potsdam BabyLab subject pool, or from a daycare centre in central Berlin. None of these participants had taken part in Experiments 4.1 or 4.2. An additional 5 participants were excluded due to poor attention (4 participants) or chance performance in the active condition (1 participant). There were no differences in age between Experiments 4.2 and 4.3 or 4.1 and 4.3 (all $p > .5$).

4.6.3 *Results*

Analysis for act-out data were conducted exactly as in Experiments 4.1 and 4.2. As there were no hypotheses concerning children’s eye movements in this experiment, and because the results of eye movement analyses in Experiments 4.1 and 4.2 on the

whole did not contribute significant conclusions to the overall interpretation of the experimental results, these data were collected in Experiment 4.3 but not analysed.

4.6.3.1 Behavioural data

Mean errors by condition and error type are reported in Table 4.13. In both active and passive conditions, participants scored at above-chance accuracy levels (chance again estimated conservatively at 50%; all p values $< .001^{***}$).

Table 4.13: Behavioural results, Exp. 4.3

Condition	% Correct (SE)	% Reversed (SE)	% Other Incorrect (SE)
Post-Subject Active	95 (2)	3 (1)	2 (1)
Post-Subject Passive	88 (3)	8 (2)	4 (1.5)

A generalised mixed-effects model with participants and items listed as random intercepts showed a performance contrast by VOICE ($z = -2.7$, $p < .05^*$), with significantly lower performance in passive conditions. A further generalised mixed-effects model, also with participants and items listed as random intercepts, confirmed that participants' incidence of committing a role-reversal error increased in passives ($z = 2.36$, $p < .05^*$). A final generalised mixed-effects model (random intercepts: participants and items) including fixed factors VOICE and AGE (in months) yielded a statistical effect of participants' age on their act-out accuracy ($p < .001^{***}$, $z = -11.74$). However, closer analysis showed that this effect rested on the performance of two children at the higher end of the age spectrum, who both performed poorly across

all conditions, perhaps partly because (unlike other participants) they were distracted by repeating every stimulus before an action. When these two participants were removed from the sample, the age effect disappeared ($p > .08$, $z = .22$), although the effect of VOICE remained ($p < .05$, $z = -2.35$). Hedges's g was calculated over children's average accuracy by condition, yielding an effect size of .58.

4.6.4 *Discussion*

Overall, results for Experiment 4.3 showed high performance across both conditions (97% accuracy in actives vs. 87% accuracy in passives). Despite the significant contrast in performance between the two conditions, children's accuracy in passives is still comparable to their performance in actives in Experiment 4.1 (pre-subject actives: 88%, post-subject actives: 90%). Given that the variance in errors for Experiment 4.3 actives is very small (based on only 4 errors for 135 data points), it would appear that children's near-perfect performance in this condition is driving the statistical significance of the effect between the two conditions, as well as the effect sizes. In addition, the reduction in children's performance on passives, while statistically significant, is a reduction of only 10%, compared to the contrast in Experiment 4.2, where children's accuracy dropped by roughly one-third between actives and passives.

Experiment 4.3 was designed to test the possibility that children's comprehension is impacted by the use of conjoined verbs or the omission of object pronouns (or the by-phrase in passives), two hypotheses that emerged from the design

and results of Experiment 4.2. On the contrary, children's overall comprehension performance was high under these circumstances. We take these results as confirmation that the performance drop in passives in Experiment 4.2 is a consequence of the main experimental manipulation, the postponement of voice information until after the lexical verb.

4.7 General Discussion

Our series of three studies investigated the developing parser's ability to assign argument roles in on-line sentence comprehension through the lens of German five-year-olds' comprehension of passives. Our line of inquiry focused on the subject-as-agent commitments that the parser makes in on-line comprehension. Prior literature has assumed that commitments to the subject as the agent (in a *seal-agent* format) are made when the comprehender encounters the subject. Based on this assumption, we manipulated whether comprehenders had access to voice information at the subject (Experiment 4.1). The subject-based hypothesis predicts that providing voice cues prior to the subject would allow comprehenders to bypass subject-as-agent misinterpretations, leading to high comprehension accuracy in passives, but lead them to a persistent subject-as-agent misinterpretation when this information was provided after the subject. The results of Experiment 4.1 showed that children performed well in passive comprehension even when voice information followed the subject. This suggested that German children's argument role commitments in the *seal-agent*

format could be successfully revised in most cases. In Experiment 4.2, we tested the alternative hypothesis that argument role commitments only become hard to revise when combined with lexical verb information into a *seal-eater* format. We found that German five-year-olds drawn from the same population as Experiment 4.1 exhibited poor passive comprehension when voice information followed the lexical verb. Experiment 4.3 controlled for possible differences in experimental parameters between Experiments 4.1 and 4.2, but found that these parameter differences (the absence of pronouns and the use of conjoined verb phrases in Experiment 4.2) could not account for the contrast in children's comprehension accuracy in Experiment 4.2.

Our results are striking in that the same test population – German five-year-olds – performed at ceiling in passive comprehension when the cue to voice was provided prior to the main verb (Experiment 4.1, Experiment 4.3), but markedly worse when this information was provided after the main verb (Experiment 4.2). Our explanation of these phenomena will revolve around the question of whether qualitative differences in argument role commitments (*seal-agent* vs. *seal-eater*) could result in a contrast in children's ability to revise misinterpretations, and the reasons why information from lexical verbs might be especially powerful in locking the developing parser into a strong commitment that is difficult to revise. Section 4.7.1 discusses the value of using eye movement data as an index of argument role assignment in online processing. Section 4.7.2 highlights several possible explanations, based on conclusions drawn in previous investigations of children's passive comprehension, that we do not believe are capable of explaining the present pattern of results. Section 4.7.3 lays out a processing-based explanation of these

results, and Section 4.7.4 situates our findings within a broader context of experimental outcomes on children's comprehension of passives.

4.7.1 *Eye movements as an indication of argument role assignment*

In Experiment 4.1, neither cue timing (providing the cue to voice before or after the subject) nor voice (active or passive) yielded contrasts between children's looks to targets and distractors following the onset of the auxiliary. Similarly, no contrast emerged during the pronoun region, with the exception of a single marginally significant time bin. This indicates that regardless of voice and cue timing, children's argument role commitments did not lead to a difference in eye gazes at the earliest stage at which they could have generated looks towards an appropriate target. It is possible that more information was required in order for any eye movements to be generated, and that once the second event participant was mentioned, children were able to compile the linguistic information they had received. However, this does not seem to have led to the generation of looks in the direction of appropriate targets within the regions of interest.

In Experiment 4.2, we saw the expected pattern of eye gazes towards the end of the auxiliary region, where children's fixation preferences indicated a clear differentiation between a preference for the appropriate target by 600 ms of the onset of the auxiliary. This timing is in keeping with the typical delays between encountering linguistic information in the input and generating eye movements on that basis, which is around 200 ms in adults (Matin et al., 1993) and 400 ms in

children (Huang et al., 2017). A surprising pattern of data emerged in the verb region, where children's eye gazes showed a fixation preference for the correct target (fish for actives, shark for passives) 400 ms prior to the onset of the auxiliary, which is the only indicator of voice in the stimulus. One possibility is that children are exploiting some form of coarticulatory information early in the sentence, prior to the voice cue. The past participles for the majority of the verbs used in this experiment ended in [ən], whose coronal nasal may take on the labial phonetic qualities of the adjoining initial [v] of the auxiliary *wurde*. However, this explanation seems somewhat unlikely, given that this coarticulation would appear only in the final syllable of the verb region, yet the contrast in fixation preferences emerges earlier.

It is nonetheless clear that German learners have a strong subject-as-agent bias, as it surfaces in the behavioural data, with a majority of the incorrect responses in the passive condition in Experiments 4.2 and 4.3 due to role reversal errors (persistent subject-as-agent misunderstanding). Given other published data, it seems that subject-as-agent biases on their own may simply not generate reliable contrasts in eye movement data. For instance, in the analogous condition in Huang et al. (2013, "expressed NP" condition in which the voice marker follows the subject), the ratio of looks to fish-type objects and shark-type objects did not differ during the initial NP or the following voice cue region. In English, Huang et al. (2017) do not report eye movement data prior to the onset of voice morphology (*the seal was eating...* vs. *...eaten by it*), but visual inspection of the proportions of looks towards seal-type objects, fish-type objects and shark-type objects suggests that these may be fairly similar. Note that the Huang et al. (2017) statistical analysis took into account gazes

from 400 ms after the onset of the voice cue and collapsed gazes from each participant and condition in a 1000-ms time window into a single data point. This analysis method effectively evaluates only gazes from the very end of the sentence, when comprehenders have full access to all information. It is therefore difficult to say conclusively whether the differences in fixation preference that emerge after the onset of voice morphology in Huang et al. (2017) are due specifically to voice information becoming available, or whether this contrast reflects the outcome of multiple overlapping processes that take into account information from different parts of the sentence. Kukona et al. (2011, Exp. 2) contrast adults' eye movements while hearing sentences like *Toby was arrested/noticed by the policeman*. While this study was designed to measure the predictive impact of verb information on anticipatory looks towards likely agents (*policeman*) or likely patients (*crook*), the eye movement data from the non-predictive condition (*notice*) with a competitor in the visual world display (both *policeman* and *crook* visually represented) most closely match the challenging conditions of passive comprehension in our experiments. In this condition, Kukona et al.'s (2011) data show that adult eye movements do not clearly differentiate between looks to likely agents and patients until the final noun region, i.e. when the likely agent (*policeman*) is mentioned in the stimulus. These results from Kukona et al. (2011) and Huang et al. (2017) highlight the difficulty of clearly attributing contrasts in eye movement data to cues to voice. Given this challenge, we rest the remainder of our discussion on our behavioural results.

4.7.2 *Alternative explanations: frequency, flexible word order, deficient linguistic representations, differences in experimental paradigms*

In this section, we briefly explore some explanations for children's difficulty with passives that have been put forward in prior work. Specifically, we investigate the ability of explanations based on frequency, flexible word order, deficient linguistic representations, or differences in our experimental paradigms to explain the pattern of results we observe between Experiments 4.1, 4.2 and 4.3.

Some studies have tracked the frequency of passives in the child's linguistic input and found a higher incidence of passives in learners' production when input frequency is high (Alcock, Rimba, & Newton, 2012; Allen & Crago, 1996; Demuth, 1990; Demuth et al., 2010; Pye, 1991; Pye & Poz, 1988). This appears to affect passive production, but also comprehension: Sesotho learners reach 70-75% passive comprehension accuracy by age 3 (Demuth, Moloji & Machobane 2010), whereas English learners do not reach this level of accuracy until two or three years later (Turner & Rommetveit 1967, Maratsos et al. 1985). If children are more frequently exposed to passives, the argument goes, they get more experience parsing passives, and ultimately attain high levels of comprehension accuracy in passives earlier than their peers learning other languages. Corpus work (by us and others, see section 4.3 and references in Huang et al., 2013) suggests, however, that German passive frequencies are comparable to frequencies in English and Mandarin, in which children display marked difficulty with passive comprehension at the same age as our German participants showed high comprehension in Experiments 4.1 and 4.3. Most

saliently, if it were the case that German input to the learner somehow favoured the acquisition of passives, this could still not explain why children's performance reflected a high level of proficiency with passives in Experiments 4.1 and 4.3, but not Experiment 4.2.

Some prior work demonstrating that German children reach certain performance benchmarks at earlier ages than peers learning English suggests that the flexible word order of German may provide learners with an advantage in reanalysing subjects as patients. Ashermann, Gülzow & Wendt (2004) argue that German learners master passives earlier than English learners because German provides reliable cues to syntactic structure beyond word order. In particular, the authors argue that subject-verb agreement is a highly reliable indicator of subjecthood in German, setting it apart from morphologically impoverished languages like English, and perhaps contributing to children's ability to parse sentences with variable word order.

This logic is insufficient for several reasons. Chief among them is the fact that the difficulty with passive comprehension is not about identifying subjects, but about identifying agents, which, contrary to argument role assignment in active sentences, do not align with syntactic subjects. This argumentation predicts that German children should have as much difficulty as English children, if difficulty in passive comprehension arises as a function of subject identification, no matter what the original cue (word order in English, subject-verb agreement in German). Yet if German children's putative facility with word order flexibility indeed led to success in passive comprehension, the differences in word order between Experiments 4.1 and 4.2 should not have resulted in such a vast gap in performance on passive

performance. Furthermore, as noted in Section 4.3, although German word order allows object-initial constructions, these are rare. It is therefore implausible that experience with flexible word order would lead German learners to greater success rates in parsing non-canonical constructions, all the more so when the main challenge presented by passives is non-canonical argument role assignment rather than non-canonical word order per se.

We note that there is one way in which word order flexibility might account for German children's higher performance in passive comprehension. Phillips and Ehrenhofer (2015) argue that children may be able to acquire low-frequency syntactic constructions by capitalising on prediction error. When children parse linguistic input, they predict upcoming material (though the reliability and speed with which they do this are debated). In the case of hard-to-parse, temporarily ambiguous syntactic structures, prediction error yields negative evidence that can be used to inform learning. Our experiments show that German learners are highly capable of correctly interpreting matrix passives, but do worse in embedded passives. Children are likely to misinterpret passives they encounter in the input if these occur in embedded clauses. However, any mismatch between their interpretation and the intended interpretation would allow them to adjust the weights on the subject-as-agent interpretation of sentences containing the auxiliary *wurde* (which, as outlined in the corpus studies discussed in section 4.3, is a more reliable cue to the passive in the past tense than in the present). According to the cue-based account of children's difficulty with passives, comprehension failures are a function of children's poor ability to recognise and use cues to the passive, due to the low reliability of the form-meaning

mapping of cues and underlying representations. Using prediction error from embedded clauses that turn out to be passives to adjust weights on *wurde* such that it is less biased towards an active interpretation may boost German children's passive acquisition and therefore, lead to better performance outcomes at earlier ages.

In order to use cues to voice in order to correctly assign argument roles in online sentence processing, children must first have acquired the ability to underlyingly represent the required syntactic operations. Representational accounts hold that deficiencies in linguistic representation are the source of children's difficulty in passive comprehension. However, if the representations themselves are at fault, this would predict equal difficulty across all word orders, which is not the pattern of results we observed in our three experiments. Any account of the performance contrast we observe must therefore address the reasons why children's ability to access passive representations would vary, rather than attempt to anchor the contrast in comprehension outcomes as a contrast in children's linguistic representations themselves.

Finally, we might explain the contrast in experimental outcomes as a result of differences in experimental parameters. We see this as unlikely, however. Experiment 4.1 used long passives, whereas Experiment 4.2 used short passives. The literature is equivocal about the effect this would have on children's comprehension (see section 4.5), though if any consensus emerges from prior experimental studies, these suggest that the contrast between long and short passives would have either a neutral impact on children's comprehension, or slightly favour their performance in short passives. Yet we saw children's performance in passive comprehension drop precipitously in

Experiment 4.2. Experiment 4.2 also departed from Experiment 4.1 in its use of conjoined verb phrases with optionally transitive verbs. As this set of verbs, as well as the structure of the verb phrase, differed from those used in Experiment 4.1, it is conceivable that this was a cause of children's poorer passive comprehension in Experiment 4.2. However, Experiment 4.3, which controlled for both the long/short passive and verb issues, showed high levels of passive comprehension, comparable to children's accuracy in Experiment 1, suggesting that these factors had a negligible, if any, effect on children's passive parsing.

There is room for speculation on the cause of the differences in accuracy between Experiments 4.1 and 4.3. In particular, why did children not perform at ceiling in Experiment 4.1 actives? Results from the present studies do not allow us to provide a definitive answer, but we suspect that pronoun resolution may have contributed to the difficulty. Past work has shown that children have difficulties with anaphor resolution, overwhelmingly interpreting pronouns as reflexive (Sekerina, Stromswold, & Hestvik, 2004). In the context of the present study, children faced a difficult task of interpreting the sentence, assigning argument roles, and selecting a pronoun referent from two toys which, by design, were equally likely to occur in an event with the expressed noun (i.e. an eating event with a seal is as likely to involve a shark as a fish). The overall difficulty of pronoun resolution, paired with the need to select a referent, may have reduced children's accuracy across the board in Experiment 4.1. Nonetheless, the experimental manipulation of placing voice information before or after the subject made no statistical difference to children's accuracy in Experiment 4.1, and children continued to perform well when voice

information was provided prior to the main verb (but without a pronoun to resolve) in Experiment 4.3. We conclude that the order of voice information relative to the main verb, not the subject, is a key factor in determining children's ability to comprehend passives.

One remaining possible factor is sample differences in the participant population. Care was taken during data collection to ensure that children in the different experimental samples did not differ in any dimensions such as age, and statistical analyses of behavioural data within each data set did not reveal any effect of age. The setting of our testing locations differed, with some children tested in daycare centres and others recruited through a university baby lab, and we did not collect data on children's socio-economic status, which has been shown to influence children's comprehension of passives (Huang et al. 2017). However, the difference in average monthly net incomes between test locations is at most €100 (Amt für Statistik Berlin-Brandenburg, 2015; Potsdam, 2015), which hardly constitutes a meaningful difference in socio-economic status.

4.7.3 *Argument roles, verb information, parsing and commitments*

The results of the present studies demonstrate that German-speaking children's ability to act on correct interpretations of passive sentences depends crucially on when they receive information about voice. Here we explore what it is about lexical verbs that might make them pivotal in children's comprehension of passives.

At root, we assume (with Huang et al. 2013, Huang et al. 2017) that children's poor comprehension is an issue of revision failure after incorrectly

assigning an agent role to the subject. In prior literature, this misassignment has been implicitly assumed to take place at the subject itself. However, our results suggest that this is an oversimplification. Subject-as-agent misinterpretation did not have a lasting negative impact on children's comprehension when voice information was presented prior to the lexical verb, suggesting that any commitments to a subject-as-agent interpretation that are made at the subject, in the *seal-Agent* format, are easily overcome, whereas a commitment of the form *seal-Eater*, which combines argument role information with information from the lexical verb, is difficult for children to revise.

The comprehender's ultimate ability to correctly interpret a sentence relies on the interplay of commitments the parser makes as comprehension unfolds, and the comprehender's ability to back off from such commitments if later-arriving information suggests that an initial parse was incorrect. By "commitment," we mean the evolving intermediate parse that is built on-the-fly, on the basis of incomplete information, and whose construction is determined by bottom-up information, for example the presence or absence of a complementiser in relative clauses in English, but also shaped by statistical or representational properties of the language, such as a subject-as-agent parsing bias. There has been a great deal of discussion concerning whether, in the face of temporarily ambiguous bottom-up input, parsers construct multiple compatible parses simultaneously, or whether the parser pursues only one interpretation at a time, based on statistical input (see, among many others, Frazier & Clifton, 1996 for detailed discussion). For present purposes, we will set the details of this debate aside, but note that regardless of whether an incorrect initial parse is

constructed at the same time as the correct one or not, success in resolving it requires the comprehender to recognise input cues that indicate that the current parse is incorrect, reject the current parse, and reinterpret prior bottom-up information to yield an overall correct parse. When we speak of “revision,” we mean the rejection and reinterpretation stages of dealing with an incorrect initial parse. In explaining the present results, we focus on how variability in the nature of comprehenders’ initial subject-as-agent commitments (*seal-Agent* vs. *seal-Eater*) would lead to variability in comprehenders’ ability to revise these commitments.

The adult processing literature demonstrates that the parser’s interpretive commitments, rather than being all-or-nothing, exist on a spectrum. Adult comprehenders can easily overcome some types of initial misinterpretation, but others are more difficult. For instance, when garden-paths are caused by sentence complement ambiguities, this is less disruptive than garden-path sentences involving closure ambiguities (see discussion in Sturt, 2007). The child processing literature has yielded a wealth of evidence that has been interpreted as evidence that the developing parser is a ballistic parser, committing heavily to initial interpretations and having great difficulty in revising these (Choi & Trueswell, 2010; Trueswell et al., 1999; Weighall, 2008). This was largely interpreted as evidence that children’s revision skills are underdeveloped, possibly due to their underdeveloped cognitive control skills. Huang et al. (2013, 2017) show that speakers of Mandarin and English are able to revise a subject-as-agent misinterpretation of a passive more easily when the subject is a pronoun than when it is a full NP. These results suggest that for children,

like for adults, the nature of the commitment determines the extent to which revision is successful.

There is circumstantial evidence to suggest that verbs are especially powerful in informing comprehenders' initial commitments. Evidence from eye-tracking shows that information hard-coded in the verb, such as selection restrictions, determines comprehenders' prediction of upcoming sentence material (Altmann & Kamide, 1999, Kamide, Altmann & Haywood, 2003). It is conceivable that the status of argument role information would differ in intermediate parses built with or without lexical verbs. Adult sentence processing studies show evidence that argument role information that is extracted from cues like word order and case, but not combined with verb information, is difficult to use in prediction at short latencies (Chow, Lau, Wang, & Phillips, 2018). This inability to use argument role information in the *seal-Agent* format in the initial stages of predicting an upcoming verb has been interpreted as the reflection of a format mismatch between the input to prediction (subject-as-agent) and the desired output (verbs that this subject would be a good agent for; Chow, Momma, Smith, Lau, & Phillips, 2016). We would like to raise a complementary possibility: there may be a qualitative contrast between the argument role commitments that comprehenders can make, depending on the availability or absence of main verb information. Due to children's underdeveloped cognitive control skills, they may be able to succeed in revising pre-verbal argument role commitments, but have greater difficulty revising argument role commitments that include verb information.

We can think of at least four reasons why this might be the case (and these are not mutually exclusive). The first option suggests that the parser is highly influenced by probabilities. While cues such as word order and case provide important clues as to the likelihood of an NP being a subject or object, they are not always perfect clues: e.g. in English object-relative clauses, the initial NP is ambiguous between subject and object, and this ambiguity may lead to processing difficulties; in German, due to syncretism in the morphological paradigm and flexible word order, case cannot distinguish between subjects and objects in 30% of sentences (MacWhinney, Bates & Kliegl 1984), again leading to processing difficulty. Not all verbs assign agent roles to the subject, and not all sentences are actives. In the absence of this level of certainty provided by the lexical verb, comprehenders' commitments to a subject-as-agent interpretation may be weak enough that even children's poorly-developed cognitive control skills can overcome the incorrect parse when the cue to voice is provided prior to the lexical verb. By contrast, if a cue to voice is presented after the lexical verb, the comprehender may have committed strongly to a subject-as-agent interpretation on the basis of argument role information provided by the verb, and this commitment may be difficult for a comprehender with limited cognitive control skills to revise when voice information arrives after the lexical verb. While this account can toggle between weak and strong, easy-to-revise and hard-to-revise commitments, it does not provide a principled reason as to why a commitment that is based on verb information would be more difficult to revise.

An alternative account is based on the idea that as relational information is attached to concrete lexical concepts, it may become more difficult to revise out of

commitments built on this basis. This may be behind Huang et al.'s (2013, 2017) finding that children are more likely to succeed in interpreting passives when the initial subject is a pronoun rather than a full NP. The commitment comprehenders build in the pronoun-initial condition includes a commitment to that pronoun as a subject and agent, but the identity of the pronoun referent can only be inferred on the basis of event information drawn from the lexical verb. Similarly, the assignment of argument roles may be limited in its usefulness in comprehension if it is unknown which event the argument role belongs to: that is, *seal-Agent* may be a less concrete commitment than *seal-Eater*, and therefore easier for comprehenders with cognitive control limitations to overcome. This is consistent with the pattern of results in Experiments 4.1 and 4.2.

Our findings may also indicate that verbs are crucial to the belief updating process that allows comprehenders to move from a parse of the input to a conceptual interpretation of that parse. By “conceptual interpretation,” we mean the level of interpretation at which comprehenders are able to engage in explicit reasoning about the propositional content of the utterance, which is required in order to complete tasks like answering comprehension questions or selecting toys from a display to act out the sentence. One goal of understanding a sentence is to infer the underlying event it describes (Kuperberg, 2015; Kuperberg & Jaeger, 2016). It is possible that an intermediate parse is passed on to the conceptual interpretation process once that event is provided, presumably once the lexical verb has been encountered in the input. This would align children’s lasting garden-path misinterpretations with the lingering misinterpretations adults entertain even after encountering disambiguating

information in the input. Christianson, Hollingworth, Halliwell, & Ferreira (2001) found that adults reading garden-path sentences of the type *While the man hunted the deer ran into the woods* were likely to incorrectly answer a comprehension question based on the initial garden-path parse (“Did the man hunt the deer?”). Slattery, Sturt, Christianson, Yoshida, & Ferreira (2013) suggest that the persistence of the misinterpretation is a result of difficulty erasing the interpretation of the initial misparse. It may be the case that parses are conceptually interpreted as soon as the comprehender encounters the main verb in the input, and that children have particular difficulty mediating between competing conceptual interpretations that arise from the initial subject-as-agent parse as well as a revised subject-as-patient parse. However, if intermediate sentence parses are not conceptually interpreted until a verb is present, revision of an intermediate parse may be seamless even in comprehenders with limited cognitive control skills.

It is possible that argument role commitments, like other commitments, become harder to revise as time unfolds, and as the comprehender encounters further information that is consistent with a subject-as-agent interpretation. In Experiment 4.2, all information up to the final auxiliary was consistent with, though not completely determinative of, an active interpretation. This pattern of results is also found in Christianson et al. (2001), who found that extending the duration of participants’ garden-path parse by modifying the ambiguous NP (*While Harry chewed the steak that was brown and juicy fell to the floor*) resulted in even higher rates of lingering misinterpretation.

4.7.4 *Similar results in prior work*

While empirical results do not allow us to disambiguate between these disparate theories of parsing and interpretation, we note that our results are consistent with prior findings. Trueswell et al. (2012) showed that young speakers of Kannada were less likely than learners of Tagalog to correctly interpret argument roles. The authors explain this as a function of the placement of causative morphology relative to the “causer” NP: in Kannada, a verb-final language, causative morphology follows the NP; in Tagalog, a verb-initial language, causative morphology precedes it. We note that the contrast in children’s performance across these two languages could also be due to the relative order of causative morphology and the verb: in Kannada, an intermediate parse of NP-as-causee may be interpreted as soon as comprehenders encounter the verb, and the following causative morphology may result in a competing but ultimately unsuccessful intermediate parse. Similarly, it is possible that English-speaking children’s difficulty with interpreting passives is an outcome of the fact that disambiguating passive morphology (*was V-ed [by]* vs. *was V-ing*) appears after the verb.

We note one difference between our findings and the previous literature. In Huang et al.’s (2013) Mandarin results, similarly-aged children performed poorly overall in passive comprehension, even though voice information precedes the verb in the Mandarin SOV sentences used as stimuli in this study. Under the hypothesis advanced to explain our data, Mandarin children should have performed well. However, a variety of factors make a direct comparison difficult. Firstly, Huang et al. (2013) aimed to investigate the strength of children’s commitments based on the

identity of the initial NP, rather than the relation of voice information to the verb. The experiment did not include a condition comparable to those in our Experiment 4.2 (as it is not possible to place the Mandarin cue to passive, the co-verb *bei*, after the main verb). It may be possible that Mandarin-speaking children's mixed success (ca. 60%) in passive comprehension in the conditions in Huang et al. (2013) is in fact at ceiling: we note that by comparison, adults' accuracy in passive comprehension was also fairly low (ca. 80%), which suggests that the task of comprehending passives was difficult even for mature comprehenders. In the Mandarin-speaking children, this difficulty may have been compounded by the task of pronoun resolution, which may have been a factor in German-speaking children's lower comprehension across all conditions in Experiment 4.1.

4.8 Conclusion

Our experiments reveal a surprising variability in German-speaking five-year-olds' comprehension of passives. When cues to voice were provided prior to the main verb, children's accuracy in a challenging act-out task was high; it dropped precipitously when the cue to voice was delayed until after the verb. Comprehenders' eye gazes revealed that they were able to integrate cues to argument role information quickly and accurately, although this tracked with high comprehension outcomes in only one experiment. Overall, these results suggest that there may be a qualitative contrast between argument role commitments, depending on whether comprehenders' parses tether this information to lexical verbs or not, which correlates with their ability to

revise these commitments when provided with evidence to the contrary. These findings provide a more fine-grained view of children's argument role processing, suggesting that their difficulty with comprehending passives is not monolithic as previously thought.

5 Prediction using pre-verbal and post-verbal argument role information

5.1 Introduction

Sentence processing is rapid and incremental: as comprehenders encounter bottom-up information, the parser swiftly integrates it into a linguistic structure that is updated and refined with every incoming word, and partially guided by top-down processing biases. Identifying who does what to whom within a sentence is a key aspect of comprehending contrasts like the one between *the bull gored the cowboy*, *the cowboy was gored by the bull*, and *the bull that the cowboy was gored by*. Each of these sentences describes a goring event in which the bull is the agent and the cowboy is the theme. However, the NPs *cowboy* and *bull* differ in their syntactic positions, as well as in their order relative to the main verb. A host of prior findings (discussed in Section 5.2) has established that argument role information extracted from NPs and word order is difficult to use accurately in real-time prediction of upcoming verbs, as indexed by a lack of N400 contrast on verbs in canonical and role-reversed sentence contexts (... *which cowboy the bull had gored* vs. *which bull the cowboy had gored*). We explore the reasons behind the N400's insensitivity to argument role reversals by probing the predictions of two contrasting accounts, one based on probability distributions in lexical associations between nouns and verbs (Kuperberg 2015; Kuperberg & Jaeger 2016), and the other based on a two-step prediction generation process in which the second, role-specific process is slowed down due to a mismatch

between the format of the search probe and verbs in linguistic memory (Chow et al., 2016; Chow, Smith, Lau, & Phillips, 2015; Momma, 2016). The two accounts differ in the predictions they make about the importance of lexical categories in generating online predictions. We contrasted these accounts in an EEG experiment in which participants read English sentences with canonical or role-reversed indirect wh-questions with noun-noun-verb word order (NNV; ... *which cowboy the bull had gored* vs. # *which bull the cowboy had gored*) and canonical or role-reversed subject-relative clauses with noun-verb-noun word order (NVN; *which jockey had raced the horse* vs. # *which horse had raced the jockey*). As per the predictions of the format-mismatch account, we found an N400 contrast on the final noun in NVN contexts, but unlike previous findings, our manipulation also yielded an N400 contrast on the final verb in NNV contexts. To our knowledge, this is the first evidence that comprehenders can accurately extract and use argument role information to predict upcoming verbs at short processing latencies.

5.2 “Semantic attraction”-based accounts of role reversal blindness

A wealth of studies has found that role-reversed sentences, in which the role assignment on context nouns is anomalous but only revealed through the identity of the target verb, do not engender N400 contrasts. For instance, Hoeks et al. (2004) found no contrast in N400 amplitudes between the target verbs in Dutch sentences like *De speer werd door de atleten geworpen* (“the javelin was thrown by the athletes”) and *De speer heeft de atleten geworpen* (“the javelin threw the athletes”).

Much of the early discussion of these findings focused on two intersecting debates: the in(ter)dependence of syntactic and semantic processing, and “good enough” parsing. Kim & Osterhout (2005) placed their findings within this discussion. Their experiment contrasted active and passive sentences such that the subject was either a good recipient of an agent role from the verb or not (*The hearty meal was devoured* vs. ...*devouring...*), and measurement on the verb revealed a lack of N400 contrast, although there was a P600 contrast. Their discussion focused on the possibility of a “semantic P600:” the idea that the parser builds semantic and syntactic representations largely independently of each other, and that the semantic parse dominates in the case of conflict. According to this view, the parser constructs a semantic parse, independent of bottom-up syntactic information, in which *hearty meal* is the most plausible theme of *devour*, and due to this “semantic attraction,” there is no semantic conflict to reflect in a contrast in N400 amplitudes. By contrast, progressive morphology on the verb mismatches the semantically plausible interpretation that has already been constructed, hence the reflection of an anomaly in a P600 contrast. Kim & Osterhout (2005) bolstered this account with a follow-up experiment in which a “semantically non-attractive” subject, *the dusty tabletops*, did generate an N400 contrast at the verb. Other studies confirm this pattern in which role-reversed sentences generated a P600 contrast, but no N400. Kuperberg, Sitnikova, Caplan, & Holcomb (2003) contrasted sentences whose subjects were related to the verb, but were either a good thematic fit for the verb (“for breakfast, the boys would only eat...”) or a bad thematic fit (“for breakfast, the eggs would only eat...”), and again found no N400 contrast on the verb. Kolk, Chwilla, Van Herten, &

Oor (2003) further refined these results by contrasting semantically attractive sentences in Dutch (*the fox-nom that at the poachers-acc stalked*) with sentences that contained selectional restriction violations (*the trees-nom that in the park-dat played*). Their results revealed a P600 contrast in the semantic attraction conditions, but an N400 contrast in the selectional restriction conditions. A follow-up study used the same experimental materials (van Herten, Chwilla, & Kolk, 2006) but fully crossed the selectional restriction and role reversal anomalies. They found P600 contrasts in both sets of conditions, but the N400 contrast again only surfaced in the selectional restriction conditions. Furthermore, the N400's blindness to role reversals is not due to overall association between arguments. Chow et al. (2015, Experiment 2) used embedded indirect wh-questions with either both subject and object nouns as plausible participants in the event described by the verb (... *which tenant the landlord had evicted...*), or in which only the subject noun was a plausible event participant (... *which realtor the landlord had evicted...*). They interpreted the resulting N400 contrast as evidence that prediction is sensitive to argumenthood. Data from Chapman, Tanenhaus & Garnsey (1989) arguably show a similar pattern. Their stimuli also exploited the filler-gap dependency in embedded indirect wh-questions to investigate the effect of plausible or implausible object nouns on verb prediction (... *which customer/article the secretary had called...*) and found an N400 contrast. While their implausible object condition was not strictly controlled for selection violations or association, this study alongside several others shows that prediction differentiates between arguments and non-arguments of the verb, but that finer-

grained distinctions relying on the specific argument roles assigned to each preceding noun have less influence on immediate N400 amplitudes.

Animacy does not appear to be a factor in the absence of an N400 contrast in the comparison of canonical and role-reversed sentences. In many early studies, animacy was either used either systematically and explicitly to change the thematic fit between the subject NP and the verb (Kuperberg et al. 2003, Hoeks et al. 2004) or not controlled (Kim & Osterhout 2005). However, subsequent study showed that potential animacy confounds are not responsible for the “semantic P600” and lack of N400 contrast. Chow & Phillips (2013) addressed the source of the “semantic P600” by investigating how animacy and semantic combinability of the nouns and verb in Mandarin SOV sentences affected ERPs. In animacy-violated sentences, the verb revealed a role-reversed relationship between the two initial nouns, with the role reversal relying on an incorrect assignment of an agent role to an inanimate noun (translation: *The student baffled the math problem*). In non-combinable sentences, the final verb did not match the preceding context (*The student hung the math problem*). The experiment yielded a complete dissociation of N400 and P600 contrasts, with P600 contrasts appearing in animacy conditions, and N400 contrasts appearing in the combinability conditions. In a follow-up experiment which controlled for animacy by using only animate nouns, however, role-reversed sentences elicited a P600 contrast, although there was still no N400 contrast. Stroud & Phillips (2012) manipulated semantic attraction in Spanish sentences whose auxiliary either biased towards a passive or progressive main verb (creating a contrast analogous to the *was devoured/devouring* contrast in Kim & Osterhout, 2005). They again found no N400

contrast across semantically and non-attractive conditions, but the stimuli nonetheless elicited a P600 contrast, including in semantically non-attractive conditions. Taken together, Stroud & Phillips (2012) and Chow & Phillips (2013) suggest that the semantic P600 may be a response to the detection of implausibility, rather than strictly about the detection of a syntactic mismatch due to a dominant interpretation shaped by semantic attraction. At the same time, the emergence of a P600 contrast in role-reversed sentences that hold animacy of the nouns constant suggests that the semantic P600 is not purely a response to animacy or selection restriction violations.

The lack of an N400 contrast in role-reversed contexts has different explanations, depending on one's view of its functional significance. Under a view in which the N400 reflects the relative ease or difficulty of integrating a specific word in the sentence context (Brown & Hagoort, 1993; Hagoort, Hald, Bastiaansen, & Petersson, 2004), the lack of an N400 contrast is blamed on "semantically attractive" combinations of nouns and verbs, in which the parser constructs the most plausible interpretation regardless of bottom-up syntactic information. However, the impossibility of pinning "semantic P600" effects to semantic attraction contexts suggests that this may not be the correct interpretation. Under the view that the N400 in fact reflects lexical-associative priming (Lau, Holcomb, & Kuperberg, 2013; Lau, Phillips, & Poeppel, 2008; Van Petten & Luka, 2012), the absence of an N400 contrast is no longer tied to the presence of a P600 in role-reversed contexts: the lack of an N400 contrast is explained as the result of associations between the nouns and verbs in context.

Finally, there have been notable exceptions to role-reversal blindness in German, Turkish and Icelandic, which do show an N400 contrast on target verbs in role-reversed sentences (Bornkessel-Schlesewsky et al., 2011; Schesewsky & Bornkessel-Schlesewsky, 2009). These authors' (Bornkessel-Schlesewsky et al., 2011; Bornkessel-Schlesewsky & Schlesewsky, 2008; Bornkessel & Schlesewsky, 2006) functional interpretation of the N400 differs from ours in several key respects. They suggest that the contrast between languages that do elicit N400 contrasts in response to role reversals (German, Turkish, Icelandic) and languages that do not (English, Dutch) is the extent to which argument role information correlates with the arguments' linear order. According to their extended Argument Dependency Model (eADM), which was developed to account for cross-linguistic variability in the presence and absence of N400 and P600 effects in role-reversed sentences, processing takes place in two stages. At the first stage, processing arguments requires a "compute prominence" step, which takes into account different sources of information (such as linear order, animacy, case) to compute argument role assignment independently of the verb itself. If the word to be processed is a verb, this requires a "compute linking" step, which determines which arguments align with which of the verb's argument roles. Additionally, "plausibility processing" operates separately at this stage on the basis of world knowledge and lexical-semantic association. According to the eADM, conflict or violation in any of the three separate processes at this first stage result in more negative N400 amplitudes. During the second stage, "generalised mapping," the output of "plausibility processing" and "compute linking"/"compute prominence" are combined, and any violations at this

stage result in more positive P600 amplitudes, resulting in “semantic P600” effects. The eADM thus explains the presence of N400 contrasts on target verbs in role-reversed sentences in German, Turkish and Icelandic as a result of conflicts in the “compute linking” or “compute prominence” processes. The additional presence or absence of a P600 effect is the result of whether that conflict can be resolved by reversing arguments. In German (Schlesewsky & Bornkessel-Schlesewsky, 2009), case information locks in the interpretation of arguments in syntactic positions: in an embedded clause like ... *dass der Schalter-NOM den Techniker-ACC bedient* (“that the switch-NOM operates the technician-ACC”), the role reversal that becomes apparent at the target verb cannot be resolved by reversing arguments due to the unambiguous case marking on the nouns, and this results in a P600 contrast relative to the canonical control. On the other hand, in an embedded clause without definite articles (... *dass Schalter-NOM-PL Techniker-ACC-PL bedienen*, “that switches-NOM operate technicians-ACC”), case information for these nouns cannot be recovered (since case is zero-marked on these specific nouns). The role reversal can be repaired by reversing the order of the arguments, eliciting no P600 contrast. It is possible that the information types that Bornkessel-Schlesewsky and Schlesewsky refer to as being required for “compute prominence” or “compute linking” have an immediate impact on prediction, but we do not believe that the N400 is a direct reflection of the processes “compute prominence” or “compute linking” themselves. For this reason, we explore the ramifications of the prediction account of the N400 in the following section.

5.3 Prediction, competition and candidate generation

Although N400 amplitude is not sensitive to argument roles at short processing latencies in languages including English and Dutch, an N400 effect does emerge at longer latencies. This is difficult to reconcile with accounts that presume that argument role information is on principle impossible to use in prediction. We instead suggest that these timing contrasts reflect delays in generating verb predictions from preceding argument information.

Chow et al. (2018) designed SOV sentences in Mandarin (where the co-verb *ba* unambiguously indicates the subject and object status of the two NPs) whose final verb cloze probability contrasted depending on word order (canonical or reversed). They found no N400 contrast when the target verb immediately followed the two NPs, but when a temporal adverb (e.g. *last week*) was inserted between the NPs and the target verb, an N400 contrast emerged for items whose average target probability in the high cloze condition exceeded 40%. Any account of the N400's systematic blindness to argument role information must therefore explain not only the N400's initial insensitivity, but also the emergence of an N400 contrast at longer processing latencies and at higher cloze probabilities.

Answering this question will at least partially revolve around an investigation of how candidates for upcoming sentence material are generated even in contexts that do not involve argument role information. The cloze completion task (Taylor, 1953) provides an offline means to probe the expectations readers generate as they complete

a sentence task, and N400 amplitude has fairly consistently been shown to track offline cloze probabilities (Kutas & Hillyard, 1980, 1984). Although cloze has been used for decades to investigate the N400, taken at face value, this measure collapses three distinct sources of contrast, and it is difficult to determine how these differentially impact N400 amplitudes. Van Petten & Luka (2012) point out that it is uncertain whether N400 amplitude reflects the degree of constraint in the sentence context (the extent to which the sentence context restricts the number of possible continuations) or the predictedness of a given word in context. A sentence might be highly constraining, such that a target with a 40% completion rate is a relatively poor completion. To adapt a well-worn example, *He mailed the letter without a stamp/an address* might have cloze values of 60% for *stamp* but 40% for *address*. Yet 40% might be the highest cloze completion in a less constraining sentence like *She left the house without her coat/shoes/umbrella* (which could have a roughly 40%/40%/20% split). In addition, the distribution of the competing completions matters. A target might have a cloze probability of 40% in context, but it might be the highest-probability completion among two others each taking up 30% of the remaining completions, or it could be the winning candidate amongst ten other completions that are each generated with a 6% probability. Should these different targets, each with a 40% cloze completion probability, be considered equivalent in terms of their predictedness? These distinctions are extremely difficult to control for in experimental investigations, and in addition, there is a host of additional factors that impact the N400, such as category membership (Federmeier & Kutas, 1999).

Nonetheless, we present a promising account by Staub (2010, 2015) which addresses the issue of the distribution of completion candidates.

Staub et al. (2015, 2010) conceptualise cloze data as representing a snapshot of the generative process which yields sentence completions. For each sentence fragment, a large group of participants generates one completion each, with each individual supplying what they consider the “best completion” for that sentence fragment. This yields a range of completions with varying distribution profiles. Staub et al. (2010, 2015) suggest that the final distribution of cloze completions is representative of the distribution of competing candidates within each individual prior to the selection of a winner. The strength of a competitor in this generation process determines its likelihood of getting selected as the final winning completion. Staub et al. (2010, 2015) conceptualise competitor strength as the speed with which a certain candidate can be generated, and demonstrate that the relative speed of generation of candidates in the “cloze race” can impact the dynamics of the generation process. However, they do not explore what would make a competitor especially strong in a given context. It is also unclear to what extent the probabilities of different competitors could dynamically change over time: although Staub et al.’s (2015) computational models help explain how the likelihoods of different competitors winning might change slightly, it is unclear whether and how candidates might be generated at different points within the cloze race, perhaps contributing to their overall lower or higher probability of being sampled as the winning competitor.

It remains to be explained why, in the case of the N400’s short-term insensitivity to argument role information, the generation process initially generates

verb candidates that are incompatible with the preceding context's argument roles, yielding a facilitated N400 for role-reversed verb targets, and how these inappropriate candidates are nonetheless suppressed enough that they do not appear at all in offline cloze completion data.

5.4 Argument roles and prediction mechanisms

Our discussion of the N400's initial blindness to role-reversal anomalies focuses on the enduring puzzle of why argument role information does not have an immediate impact on verb predictions. Role-reversed sentences give rise to P600 contrasts on the target verb even when animacy is controlled (Chow & Phillips, 2013; Chow et al., 2015), indicating that argument role information is extracted early and anomalies in argument role assignment can be detected immediately. However, the lack of an N400 contrast at short latencies (Chow et al., 2018) indicates that although comprehenders have access to argument role information, it does not immediately impact prediction.

In explaining the N400's apparent blindness to argument roles, scholars have taken two broad classes of approaches. The first approach (Kim & Osterhout, 2005 and others) assumes that the N400 is insensitive to argument role information because bottom-up information is parsed separately by its syntax and semantics, with plausibility-driven interpretation winning out over syntactic interpretation, leading to a lack of N400 contrast. This cannot be straightforwardly true given that N400 effects surface when comprehenders have more time for parsing (Chow et al., 2018). Some

accounts explain this by suggesting that argument role information can only be accessed at a delay, or that it is accessed early but used late in prediction; see Chow et al. (2016) for a discussion and rebuttal of these claims.

Two accounts take the alternative approach of locating the difficulty with using argument role information in real-time prediction at the level of using available information in order to efficiently search through memory and generate predictions for upcoming verbs. Both accounts assume that the purpose of comprehension is to infer the underlying event that is being described, but the accounts differ in terms of how they operationalise the search for that event based on the available noun and argument role information. Kuperberg (2015) and Kuperberg & Jaeger (2016) base their explanation in the probability distributions of associations between nouns and the to-be-predicted verbs. An opposing account (Chow et al., 2015; Chow et al., 2016) locates the prediction difficulty in a mismatch between the search probes that comprehenders can construct on the basis of noun and word order information, and the format of verbs in event memory, such that the search for a fitting verb is not complete by the time the actual verb is encountered in the input.

Kuperberg (2015) formalises the search for a verb matching the context as a need to infer an underlying event which would give rise to the specific surface syntax that the hearer has encountered. For Kuperberg (2015), this inference is probabilistic, based on the hearer's previous world knowledge and linguistic experience, and on the subset of bottom-up cues which is most reliable. If, given two participants (e.g. *ghost* and *villager*), one event (e.g. *haunting*) is especially frequent regardless of the exact nature of the argument role assignment, that event will receive most activation and

therefore not elicit an N400 contrast when the comprehender encounters the verb *haunt*. A single, lexical-associative mechanism for candidate generation could explain the N400's insensitivity to verbs in role-reversed sentence contexts, but only if that insensitivity remained constant over time. As Chow et al. (2018) show, however, role sensitivity emerges in the N400 when comprehenders have more time available for the search. Kuperberg (2015) invokes a second, role-dependent search process that generates only those verb candidates which fit existing argument role assignments. According to her account, what tips the scale in favour of one search over another is the "reliability" of bottom-up information: when there are strong but unidirectional associations between the two NPs (ghosts stereotypically haunt people such as villagers, while there is no strong prediction for an event in which a villager would be doing something to a ghost), the parser settles on the event generated by the lexical-association mechanism. However, when there are strong associations in both directions, the parser chooses the event that matches the existing argument role assignment. Recast in the framework of prediction as competition, in Kuperberg's view, the lack of N400 contrast is therefore reduced to the issue of one event being strongly predicted in canonical sentences, but no strong competitors emerge in reversed sentences.

By contrast, Chow et al. (2016) posit two search mechanisms: one lexical-associative parallel search mechanism, which quickly activates all possible verb associates of two given NPs without taking argument role information and ranks them by activation levels, and one role-dependent serial search mechanism which assesses the ranked verb candidates' match with the preceding NPs' argument role assignment.

The ordering relationship between these two processes gives rise to a processing delay. Yet both types of search mechanism are required due to a format mismatch between the search probe that is available to the parser (*ghost-Agent, villager-Patient*) and the format in which events are stored in episodic memory. According to this account, the winning candidate from the lexical-associative search mechanism is predicted until the outcome of the role-dependent serial search is finalised, leading to a role-insensitive N400 at short processing latencies, but a role-sensitive N400 at long latencies.

These two accounts differ centrally in terms of their predictions about lexical categories. If the prediction of upcoming sentence material is determined purely by the strength of association between past sentence material (as in the Kuperberg account), this mechanism should function similarly whether the existing input consists of two nouns (as in embedded indirect object-initial wh-questions, henceforth NNV clauses) or a noun and a verb (as in embedded indirect subject-initial wh-questions, henceforth NVN clauses). If this is true, N400s on the final noun in NVN clauses should be as insensitive to role reversals as N400s on the final verb in NNV clauses. On the other hand, if the primary factor underlying N400 insensitivity to role reversals is a format mismatch between argument role-labelled nouns and events in memory, it should be possible to overcome this difficulty by providing comprehenders with a verb early. The format mismatch account predicts an N400 contrast in role-reversals on the final noun of an NVN clause, while the N400 to the final verb in an NNV clause should remain insensitive to argument role reversals. We designed Experiment 5.1 to address these questions.

5.5 Experiment 5.1

5.5.1 *Materials*

Test materials (see Table 5.1 for examples) were based on Chow et al. (2015), and consisted of sentences containing either an embedded indirect object-indirect wh-question (noun-noun-verb; NNV) or embedded indirect subject-indirect wh-question (noun-verb-noun; NVN) embedded clause. In NNV clauses, non-canonical OSV order ensured that the two nouns' argument roles were accessible before the target verb. In NVN clauses, canonical SVO order ensured that the first noun's argument role was available prior to the target noun. Targets either matched or mismatched the preceding context's argument role assignments. This was achieved by reversing the order of the nouns (NNV: *which bull the cowboy had ridden* vs. *which cowboy the bull had ridden*; NVN: *which jockey had raced the horse* vs. *which horse had raced the jockey*).

Our experimental design departs from the previous literature in fully crossing order and the target's fit to context. Each context order was therefore paired with both a high-cloze completion which matched the argument role assignment of the preceding context, and the opposite order's completion, which in this context was low-cloze. In most NVN stimuli sets, the demands of balancing a high cloze value for a completion in one order against that same completion having a low cloze value in the opposite order required us to replace nouns; see Table 5.1 for details. Filler

sentences were designed to elicit the typical N400 contrast in response to unpredicted target items (Kutas & Hillyard, 1980, 1984). These sentences contained either noun or verb targets (N-Filler vs. V-Filler conditions). Low-cloze completions were created by scrambling high-cloze targets to produce sentences whose targets were both low-cloze and implausible.

Table 5.1: EEG Sample Stimuli

<i>NNV contexts</i>	The cattle rancher remembered...	<i>mean cloze</i>
<i>a</i> Canonical, high cloze	... which bull the cowboy had <u>ridden</u> out on the range.	36%
<i>b</i> Reversed, low cloze	... which cowboy the bull had <u>ridden</u> out on the range.	<2%
<i>c</i> Reversed, high cloze	... which cowboy the bull had <u>gored</u> out on the range.	36%
<i>d</i> Canonical, low cloze	... which bull the cowboy had <u>gored</u> out on the range.	<2%
<i>NVN contexts</i>	The horse trainer saw...	
<i>a</i> Canonical, high cloze	... which jockey had raced the <u>horse</u> across the track.	36%
<i>b</i> Reversed, low cloze	... which horse had raced the <u>jockey</u> across the track.	<2%
<i>c</i> Reversed, high cloze	... which horse had thrown the <u>jockey</u> across the track.	36%
<i>d</i> Canonical, low cloze	... which gambler* had thrown the <u>horse</u> across the track.	<2%
* In NVN stimuli, the need to ensure counterbalanced cloze values for the final noun required initial noun substitutions in conditions c and/or d. These nouns were chosen to be highly associated with the remaining nouns and verbs in the stimulus set.		
<i>Controls</i>		
Verb control	During the gold rush, prospectors <u>found/knitted</u> gold in the Rocky Mountains.	36% vs. 0%
Noun control	The environmentally friendly office recycled <u>paper/nuts</u> and plastic whenever possible.	36% vs. 0%

In cloze norming, we counted semantically highly related lemmas towards the highest-cloze member of that semantic group (e.g. 20% “eaten”, 5% “devoured”, 5% “consumed” would count as 30% “eaten”). Across filler conditions, high-cloze completions averaged 36% (N-filler SD: 13; V-filler SD: 15) and low-cloze completions averaged 0% (SD: 0). High and low-cloze conditions were averaged within NNV and NVN conditions. In NNV contexts, high-cloze conditions averaged

35% (SD: 14) and low-cloze conditions averaged 1.4% (SD: 2.8). In NVN contexts, high-cloze conditions averaged 36% (SD: 17) and low-cloze conditions averaged 2.5% (SD: 4.4).

To ensure this careful balance of high and low-cloze completions across two sets of four test stimuli each, as well as two sets of fillers, stimuli were developed in an iterative series of 19 web-based cloze norming experiments on Amazon MTurk. Experiments varied in length from 10 to 50 minutes. Participants were paid to complete sentence fragments which ended prior to the target with “the first thing that comes to mind,” and were asked to take no more than twenty seconds to complete each sentence. The final test and filler stimuli set contained completion data from 30 participants each, but the process of developing stimuli across 19 experiments required data from a total of over 900 individual participants. No participant completed more than one norming experiment (guaranteed through the use of a Unique Turker ID, <https://uniqueturker.myleott.com/>).

We created sixty stimuli of each type (NNV, NVN, N-Filler, V-Filler) and distributed them across four lists in a Latin Square design that ensured that no participant saw more than one condition for each item set. Each list contained 50% high and low-cloze items and an equal proportion of stimulus types. Item order was randomised between participants.

5.5.2 *Procedure*

Participants sat comfortably at a distance of ca. 100 cm from a monitor and read sentences that were presented in RSVP using 24-point font. Following Chow et al.

(2015), a fixation cross was shown for 500 ms at the beginning of each trial. Words were displayed for 300 ms, with a blank screen for 230 ms after each word (total SOA = 530 ms). Participants answered a yes/no plausibility question 1000 ms after the last word of each sentence (marked with a full stop). Plausibility was defined as “something that could normally happen.” To ensure greater attention, participants’ accuracy on fillers was displayed at the end of each experimental block with a note encouraging better performance where required. Participants were offered breaks after 40 stimuli or ca. every 10 minutes. Each experimental session took an average of two hours. Participants gave informed consent and were paid \$10-\$15/hour.

5.5.3 *EEG recording*

Continuous EEG measurements were collected from 29 AgCl electrodes placed on the participant’s head using an electrode cap (Electrocap International): midline: Fz, FCz, Cz, CPz, Pz, Oz; lateral: FP1, F3/4, F7/8, FC3/4, FT7/8, C3/4, T7/8, CP3/4, TP7/8, P3/4, P7/8, and O1/2. Scalp electrodes were referenced to the left mastoid online, and in an offline processing step, re-referenced to the average of both mastoids. To track eye movements, the electro-oculogram (EOG) was recorded at four bipolar electrode sites, with two electrodes above and below the left eye recording vertical EOG and a further two electrodes at the outer canthus of each eye recording horizontal EOG. Electrode impedances were below 10 k Ω for all participants (and for all but three participants, they were below 5 k Ω). EEG and EOG

recordings were amplified and digitised online at 1kHz with a bandpass filter of 0.1–100 Hz.

5.5.4 *Participants*

24 adult participants from the University of Maryland participant pool were paid for their participation (11 female, mean age 22 years, SD: 4.3). All participants were native speakers of American English, right-handed according to the Edinburgh Handedness Test (Oldfield, 1971), with corrected-to-normal vision, no reading disabilities and no history of neurological disease. Data from a further three participants were excluded for poor performance in the plausibility task, and a further five participants due to excess EEG artefacts.

5.5.5 *Results*

5.5.5.1 Behavioural results

In fillers, the plausibility judgement served as a test of participants' attention. Three participants who scored lower than 85% accuracy were therefore excluded from further behavioural and EEG analysis. For the remaining data (see **Error! Unknown switch argument.**), a linear mixed-effects model (calculated using the R package lme4, D. Bates, Maechler, Bolker, & Walker, 2015) with fixed factors CLOZE, CONTEXT and the interaction of CLOZE and CONTEXT and random factor PARTICIPANT revealed a significant contrast of plausibility ratings by CLOZE ($t = -4.1, p < .001$ ***),

but not by CONTEXT ($t = -.4, p > .5$), and with no interaction between the two factors ($t = -.55, p > .05$).

Table 5.2: Mean plausibility ratings by test condition

Context	Cloze	Mean Plausibility Rating (SD)
NNV	high	84% (.08)
NNV	low	68% (.2)
NVN	high	83% (.1)
NVN	low	64% (.2)

5.5.5.2 EEG results

Trials affected by EOG or other artefacts were removed from further analysis; this affected 15.4% of the original sample. Averages were computed separately per participant and per condition, based on a baseline of 100 ms pre-target and 1000 ms post-target. Waveforms for electrode Cz are shown in **Error! Unknown switch argument.** A repeated measures ANOVA was performed on time windows of 300-500 ms (N400 time window) and 600-800 ms (P600 window), for a mid-anterior region of interest (ROI) of eight electrodes around Cz (Pz; CPz; Cz; P3; P4; C3; C4; CP4). The ANOVA fully crossed CLOZE (high/low) and CONTEXT (NNV/NVN). For test items, in the 300-400 ms window, this two-by-two ANOVA revealed a main effect of CLOZE ($F = 4.46, p < .05^*$) but not of CONTEXT ($F = .17, p > .5$), with no interaction of CLOZE and CONTEXT ($F=.14, p > .5$). In the 600-800 ms window, this two-by-two ANOVA revealed no main effects of CLOZE ($F=1.1, p < .03$) or CONTEXT ($F = .3, p > .5$), with no interaction of CLOZE and CONTEXT ($F=.15, p > .5$).

In the control comparison, a two-by-two ANOVA fully crossed CLOZE (high/low) and CONTEXT (V-Filler/N-Filler). In the 300-500 ms window, this two-by-two ANOVA revealed a main effect of CLOZE ($F = 43.2, p < .001^{***}$) and of CONTEXT ($F = 7.4, p > .001^{***}$), with no interaction of CLOZE and CONTEXT ($F=2.6, p > .1$). In the 600-800 time window, there was no main effects of CLOZE ($F=3.9, p < .05$). There was, however, an effect of CONTEXT ($F = 28.3, p < .001^{***}$), as well as an interaction of CLOZE and CONTEXT ($F=5.1, p < .5$).

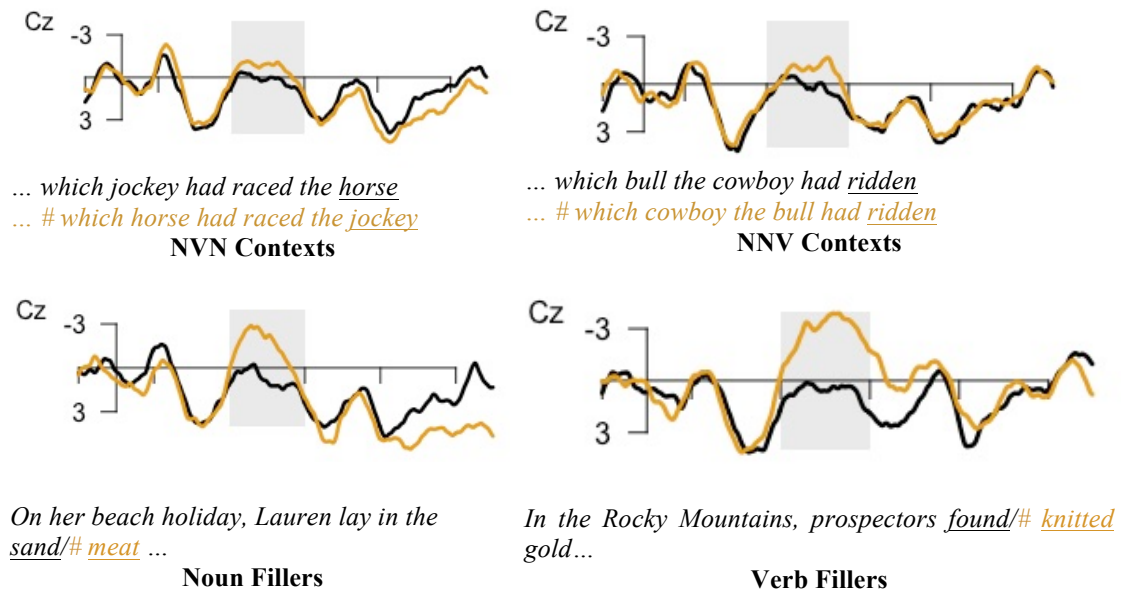


Figure 5.1: Sample EEG waveforms, Exp. 5.1.

(Shaded areas correspond to the 300-500 ms window of analysis.)

5.5.6 Discussion

A significant contrast in N400 amplitude was elicited in the manipulation of cloze through role reversals. In the absence of an interaction between cloze and context, the

statistics do not yield precise information about which of the two contexts (if not both) are the source of the cloze-based N400 contrast. Given that we did not replicate the lack of N400 contrast in role-reversed NNV contexts prevalent in the existing literature, we cannot firmly address the original research question of how lexical categories contribute to the use of argument roles in prediction. We now concentrate our discussion on two major areas: possible reasons for the difference in N400 outcomes between NNV contexts in this study, which found a contrast, and prior work, which did not; and the potential implications of differences, had any occurred, between N400 outcomes in NNV and NVN prediction contexts.

5.5.7 *Post-hoc stimulus analyses*

The NNV stimuli in the present study were intended to be designed according to the same principles as the stimuli in Chow et al. (2015). However, in light of the emergence of an N400 contrast that was not present in the prior study's results, we conducted a number of post-hoc analyses and a follow-up experiment to determine whether there were differences in stimulus make-up that could have led to the present study's unexpected and anomalous results.

Stimulus creation in this study was constrained by the need to create sets of four sentences that balanced cloze (low/high) with word order. In Chow et al. (2015), stimuli consisted of low/high cloze pairs, such that one order was paired with its high cloze verb completion, then reversed so that this completion no longer had a high cloze value. In the present study, by contrast, the additional constraint of needing to

cross word orders meant that stimuli could only be included if the high-cloze target for each order did not appear in the cloze completion data for the reversed order. This ensured that offline, each order generated a specific verb prediction, which was not at all (or barely) predicted for the reverse order. Additionally, low-cloze stimuli in the present study had to have not only a low cloze probability for the target completion, but as far as possible, this low-cloze completion had to be implausible. We fielded three efforts to quantify any difference this might have made in online prediction.

5.5.7.1 Target cloze distribution

One possible explanation for the difference in N400 outcomes between the present study and Chow et al. (2015) is that there may be differences in the target cloze values across stimulus sets. That is, the stimulus sets may differ in terms of how highly predicted the maximum cloze item is for each condition.

However, statistical investigations suggest that this is not the case. Within each stimulus set, the average maximum cloze values predicted offline for canonical and reversed sentence contexts differ (present stimuli: $t = 2.37$, $p < .05^*$; Chow et al. stimuli: $t = 3.08$, $p < .01^{**}$). A two-way ANOVA listing items as random intercepts confirmed a significant effect of cloze ($F = 14.6$, $p < .001^{***}$) but no effect of stimulus set ($F = 3.3$, $p > .05$). Any difference in the two stimulus sets can therefore not be attributed to differences in the means of the maximum cloze values across word orders in each stimulus set.

5.5.7.2 Entropy

We next took a broader view of cloze completion data, and asked whether there were differences in the probability distribution of completions beyond the probability of the most frequent completion. If the most frequent offline completion is far more frequent than any of the other completions, this might map onto a different online prediction than when the most frequent offline completion is only marginally more frequent than other completions for this sentence fragment. We computed this in the terms of Shannon (1948), who calculates entropy as the sum of each completion's probability divided over that completion's log probability. Within each stimulus set, the distribution of entropy values differed statistically between word orders (present stimuli: $t = -3.03$, $p < .01^{**}$; Chow et al. stimuli: $t = -3.6$, $p < .001^{***}$). A two-way ANOVA showed a statistical effect of word order ($F = 22.3$, $p < .001^{***}$), but none of stimulus set ($F = 2.4$, $p > .05$), suggesting that any entropy differences between high and low cloze conditions were comparable across stimulus sets.

5.5.7.3 Frequency

We next compared the log frequencies of target verbs across both experiments, based on the well-established finding that N400 amplitudes are impacted by target word frequency (Van Petten & Kutas, 1990) and the intuition that more frequent words might be generated more quickly in prediction. We identified the number of unique target verbs appearing in each condition across the two stimulus sets, and found that both stimulus sets used similar numbers of unique verbs (80 in the present study; 77

in verb targets for the stimuli used in Chow et al. (2015), including the highest-cloze completions for Chow et al.'s reversed orders, which were not used in the experiment itself). There were no statistical contrasts in the means of log frequency distribution across word orders within each stimulus set (present stimuli: $t = 2.3, p > .01$; Chow et al. stimuli: $t = -2.3, p > .01$). However, a two-way ANOVA revealed that although the contrast in log frequencies did not vary by word order ($F = .125, p > .05$), it did vary between stimulus sets ($F = 8.5, p < .01^{**}$).

5.5.7.4 Subject-verb cosine relationship

As outlined above, comparing the sets of stimuli used in Chow et al. (2015) and in the present study did not yield measurable differences between the two stimulus sets in terms of differences in entropy or cloze probability distributions, though a mild difference emerged in target frequencies. The standard assumption is that N400 amplitude is closely related to cloze probability, and yet the role-reversal literature has already established that this cannot straightforwardly be the case. One issue is that existing models comparing the offline outcome of cloze completion tasks to online prediction mechanisms (Staub et al., 2015, 2010) have yielded useful insights about the possible dynamics between competitors in the process of generating a prediction, but have not formulated clearly what characteristics would make a competitor strong (and therefore more likely to be sampled as the winning competitor) or fast (and therefore more likely to reach a certain activation threshold before other competitors in the cohort). Ettinger's (2018) measure of the subject-verb cosine relationship offers one possibility. The cosine relationship between two words provides a measure of the

extent to which they co-occur in the same linguistic contexts. Ettinger's (2018) calculations showed a similar subject-verb cosine relationship between embedded subjects and verbs in Chow et al.'s (2015) stimuli, but found that there was a greater discrepancy in similarities between the high and low cloze stimuli in the present study. In the present study, this difference is significant ($t = 2.3, p < .05^*$), whereas it is not in the Chow et al. (2015) stimuli ($t = -.3, p > .5$). It may be the case that embedded subjects take a leading role in generating verb candidates despite appearing late in the input (after the object is available). If so, a small contrast in subject-verb cosine relationship across reversed and canonical sentence contexts could map onto a similarity in predictions generated from the embedded subject, leading to a lack of N400 contrast, as seen in Chow et al. (2015). On the other hand, a difference in subject-verb cosine relationship could map onto disparate verb predictions generated on the basis of that embedded subject, leading to an N400 contrast, as seen in the present study.

5.5.7.5 Plausibility: a follow-up experiment

Finally, we attempted to quantify the difference in reversibility of the high cloze verb completion for each context order, with a view to establishing whether there was any difference in the strength of this contrast across the two stimuli sets. We hypothesised that constraints on stimulus creation in the present study might have led to reversed orders that elicited completions which did not rely on argument role information, as high-cloze completions for the reversed noun contexts were not specific to that order and therefore, not specific to that assignment of argument roles. We expected robust

plausibility rating contrasts between high and low cloze completions within each stimulus set. However, if there is an underlying difference in the degree to which order (canonical/reversed) predicts a role-specific verb, we expected this to be reflected as a between-stimulus set difference in plausibility ratings when high-cloze completions for reversed orders were applied to the canonical order.

We conducted an online plausibility rating task on Amazon Mechanical Turk, comprising a total of 60 participants. Cloze completion data from norming studies in Chow et al. (2015) were used to create an additional set of conditions, such that the resulting Chow et al. stimuli matched those used in the present study, in fully crossing not only cloze (low/high) but also word order (canonical/reversed). The resulting 120 stimulus sets were distributed across four experimental lists in a Latin square design which ensured that each participant saw only one condition per item. Participants were asked to individually rate sentences on a Likert scale ranging from 1 (very implausible) to 7 (very plausible). One participant was excluded due to illicitly repeating the task. For all participants, two item sets were excluded for technical reasons. For an additional 14 participants, a further three item sets were excluded for technical reasons. For each item in each condition, there were 13-14 ratings, spread among a total of 55 participants.

We ran a mixed-effects model to investigate the between-stimuli set contrast in plausibility ratings between reversed contexts paired with high-cloze verbs and the corresponding canonical context paired with that same verb to yield a low-cloze completion. The fixed effects were the interaction of condition and stimulus set. We also included random intercepts for item and participant. Statistical results showed an

effect of condition (high-cloze reversed order was rated as significantly more plausible than the low-cloze reversed order condition, $t = 36.6$, $p < 0.001^{***}$), an effect of stimulus set (high cloze conditions were rated more plausible than low cloze conditions in Chow et al. 2015 items than in the present study's items, $t = 3.9$, $p < 0.001^{***}$) and an interaction of condition and stimulus set ($t = -7.37$, $p < 0.001^{***}$).

We repeated this approach to compare between-item set plausibility ratings for canonical orders with high-cloze verb and the reversed order paired with that same verb to yield a low-cloze completion. The fixed effects and intercepts replicated the design of the previous model. Here, there was again an effect of condition, with low-cloze completions rated as significantly less plausible than high-cloze completions ($t = -57.7$, $p < 0.001^{***}$). However, there was no effect of stimulus set ($t = -0.268$, $p > .5$), nor was there any interaction between stimulus set and condition ($t = 0.232$, $p > .5$).

We conclude from these statistical analyses that plausibility ratings differed significantly within each stimulus set for high and low cloze completions. Across the two stimulus sets, the contrast in plausibility ratings seems equally strong when high-cloze completions in the canonical order are applied to the reversed order. However, in line with our prediction, plausibility ratings for high-cloze completions of reversed orders applied to the canonical order yield a contrast between stimulus sets. This suggests that the implausibility constraint in the present study's stimuli led to the selection of stimuli that gave rise to more distinctive, non-overlapping offline prediction profiles.

5.6 General discussion: Argument roles in prediction and retrieval

The present study aimed to explore the extent to which the lexical category of words in context influences the use of argument roles in prediction. The experiment was designed to replicate the results of Chow et al. (2015, 2018) and others, in which a comparison between canonical and role-reversed object-relative clauses elicited no N400 contrast on the final verb, which had to be predicted on the basis of two preceding nouns. We also aimed to extend these findings to a subject-relative context, in which comprehenders had to use a noun and a verb to predict a final noun. We hoped that this manipulation would help differentiate between two different accounts of the lack of N400 contrast in role-reversed sentences. Both the probability-based account (Kuperberg 2015, Kuperberg & Jaeger, 2016) and the format mismatch account (Chow et al., 2016) assume that verb candidate generation involves two mechanisms, a lexical-associative generation mechanism which does not take argument roles into account and a role-specific generation mechanism which does. The probability-based account assumes that both of these mechanisms act in parallel, but the “reliability” of bottom-up information (such as word order, which makes it possible to extract argument roles) in inferring the underlying event determines whether the verb that is ultimately predicted is the outcome of the lexical-associative or the role-specific generation mechanism. This account attributes the lack of N400 to the presumed unidirectionality of events in the canonical sentences in Chow et al. (2015), such that regardless of word order (canonical or reversed), only one event – the outcome of the lexical-associative mechanism – is predicted. Chow et al. (2016) counter this by noting that cloze testing of reversed sentence fragments yielded

maximum cloze values comparable to the maximum cloze values of targets in canonical sentences, suggesting that both orders generate distinct verb predictions offline and therefore refuting Kuperberg's (2015) point about unidirectionality.

By contrast, the format mismatch account assumes that the two mechanisms act in sequence: first, a parallel search mechanism generates all events compatible with the preceding nouns, regardless of argument role assignment; the output of this generation mechanism yields an ordered list of events ranked by activation, which must be checked by a slower serial search mechanism that verifies whether a candidate matches the argument role assignment of the preceding nouns. By hypothesis, this second process is slow because of a mismatch between the format of the search probe that is used to retrieve events from linguistic memory, and the format of those events in linguistic memory. This account thus attributes the lack of N400 contrast at early processing latencies to the prediction of only a non-role specific verb candidate up until the role-specific generation mechanism has concluded its search, yielding the timing effects observed in Chow et al. (2018).

The current results contradict both of these accounts in a number of ways. Most notably, we found a clear N400 contrast in role-reversed NNV contexts, where numerous prior studies found none. However, the lack of difference in N400 outcomes between NNV and NVN contexts provides us with no way to distinguish the predictions of the probability-based and the format mismatch account. In this discussion, we explore how both accounts do and do not explain the present results, how modifications to each account could bring it in line with the present findings, and

the consequences of any putative contrast in lexical categories for theories of the architecture of linguistic memory.

5.6.1 *How could the format mismatch account explain the present results?*

The current version of the format mismatch account is unable to explain the present results because according to that set of proposed mechanisms, the role-specific search mechanism is presumed to be slower, by design, than the lexical-association search mechanism. By contrast, in the present study, the final verb target in role-reversed NNV sentence contexts elicited more negative N400 amplitudes than the target in canonical sentences, indicating that the predictive mechanism was, in fact, able to take argument role assignment into account. In order for the format mismatch account to accommodate these findings, it would be necessary to make the lexical-association mechanism either systematically generate only one verb candidate (which is appropriate in one word order but not the other, yielding no N400 contrast) or generate two verb candidates but somehow, at this processing stage, select just one of them as the predicted item. Alternatively, the role-specific serial search mechanism would have to be revised to be able to operate at different speeds, yielding a slow narrowing-down of the candidate field to just role-appropriate candidates the stimulus set used in Chow et al. (2015), but acting quickly in the present study.

The assumption inherent in Chow et al. (2016) is that the lexical-associative mechanism generates all candidates that are consistent with either argument role assignment. That is, regardless of whether the incoming sentence fragment is *which*

customer the waitress had... or *which waitress the customer had...*, the lexical-associative mechanism should always generate both *serve* and *tip*, thus yielding no contrast in N400 amplitude on *serve* when it follows a role-reversed context. This is consistent with Chow et al.'s (2015, 2016) maximum offline cloze values of ca. 26% in canonical and 22% in reversed sentence fragments. Staub et al. (2015), following Smith & Levy (2011), suggest that cloze completion data provide an offline snapshot of the field of competitors within each participant's generation process. According to this view of prediction, each individual generates a field of competitors of varying strengths, and the final offline completion each individual settles on is the result of sampling from this distribution of candidates. The likelihood that a candidate is picked as the winning cloze completion in the offline measure is therefore conceptualised as being directly related to the strength of that competitor in the online candidate generation process. Assuming that *serve* and *tip* have equal likelihood of being sampled as the winner of the cloze competition, at short processing latencies, they should be equally predicted and therefore both elicit facilitated N400 amplitudes as the target of either context. In Chow et al. (2015), however, only one of these two predictions was used as the target, meaning that there is no way to compare the facilitation of *serve* relative to the facilitation of a competing verb prediction.

Taking the original format mismatch account at face value, the lexical-association mechanism generates an ordered list of verb candidates, which then forms the input to the role-specific mechanism. If activation levels were determined in such a way that role-appropriate verb candidates are always at the top of the list, this could yield the contrast that allows the role-specific mechanism to quickly identify a role-

appropriate verb candidate for prediction, explaining the present study's N400 contrast. One option for such a submechanism would be to suggest that prediction at this stage of processing actually reflects a combination of the conditional probabilities of the target given the context nouns, and the context nouns given the target. While such a calculation is potentially able to yield a high activation level for *serve*, it is not able to yield differential activation levels for contrasting verb candidates that would lead to the N400 contrast found in the present study, as multiplication and addition operations yield the same outcome regardless of the order in which they are performed.

At first glance, this difficulty might be ameliorated by assuming that predictions decay over time. The strength of activation for nouns given the verb might be weighted according to how much time has passed since these nouns were encountered in the input. A prediction that is generated on the basis of the embedded object might experience some decay while a second prediction is generated on the basis of the embedded subject. This account, however, has the opposite problem: while it might be able to account for the robust contrast in prediction in the present study, it is unable to account for a lack of such a prediction contrast in Chow et al. (2015), if the predictions generated by the lexical-association mechanism are indeed more sensitive to the embedded subject as a simple matter of decay over time.

A further alternative to consider is whether to adjust the function of the role-specific serial search mechanism such that, under certain circumstances, it might be made to act faster and therefore, yield role-specific verb predictions fast enough to impact prediction. This could be achieved in one of two ways: assuming a way for the

role-specific mechanism to operate earlier or later in the prediction process to yield the timing contrasts between Chow et al. (2018) and the present study; or assuming that the role-specific mechanism itself operates at varying speeds. If it is the case that the lexical-association mechanism generates both *serve* and *tip*, and that these are equally strong predictions, then both of these event predictions should be at the top of the list for the serial search mechanism to check against the preceding context's argument role assignments. It is necessary to assume that any operation of the role-specific mechanism is slow in order to yield the timing contrasts found in Chow et al. (2018), which necessarily entails that checking and rejecting as few as one verb candidate prior to moving down the list and identifying a correct item takes long enough that this mechanism's output cannot impact prediction at short latencies. If we assume that the role-specific mechanism can begin its processing earlier under specific circumstances, there would have to be some triggering factor during the lexical-association process that would allow the role-specific mechanism to get a head start in checking candidates in Chow et al. (2015)-style stimuli, but not in the present stimuli. If we assume that the role-specific mechanism is inherently variable in its processing speeds, it would be necessary to posit a reason why it would have been systematically slow in processing verb candidates in Chow et al. (2018), but not in the present study. Again, given that exploration of the possible distinctions in the cloze completion data for each stimuli set yielded no systematic difference, it seems that any contrast in the role-specific mechanism's operating speed must be caused by some underlying difference that is not reflected in offline cloze data.

In summary, updating the existing version of the format mismatch account repeatedly runs into the same difficulty: it is impossible to tweak that account's mechanisms such that they would generate verb predictions that yield an N400 contrast for the stimuli used in the present study, but no such contrast for the stimuli used in prior work. However, this account's strength lies in its ability to explain the timing contrasts found in Chow et al. (2018) and Momma (2016). The format mismatch account therefore retains a partial ability to account for the observed phenomena.

5.6.2 *How could the probability-based account explain the present results?*

The probability-based account supposes that the contrast between role-blind and role-specific prediction generation lies in the “reliability” of the bottom-up information that is being used to infer the underlying event. In Kuperberg's (2015) account, the choice between using the full set of available information (i.e. nouns including their word order, acting as proxy for argument roles) or a reduced subset of available information (nouns only, excluding word order information) is determined by which of these sets of information yields greater certainty in prediction. This assumes an asymmetry in the events that can be predicted from two nouns: for instance, the combination of *customer* and *waitress* would need to yield a higher-probability prediction than, say, the combination of *customer-AGENT* and *waitress-THEME* in order for the role-blind verb candidate to be predicted. Chow et al. (2016) rebut this proposal, noting that offline cloze completion data suggest symmetry in the prediction

probabilities of non-overlapping events generated on the basis of different word orders. If this is straightforwardly the case, the probability-based account cannot account for the data in Chow et al. (2015), even without considering the timing differences observed in Chow et al. (2018).

However, as alluded to above, the symmetry of predictions emerging from offline cloze completion data must obscure some differences in the way that these predictions are generated online, as otherwise the same pattern of results would be expected to hold across Chow et al. (2015) and the present study. An illuminating part of the discussion in Hoeks et al. (2004) suggests that offline predictions may be generated at different speeds. Hoeks et al. (2004) conducted an offline cloze norming task revealing that participants found it more difficult to come up with completions for reversed sentences, e.g. one in which *javelins* are the agent of an event that has *athletes* as its theme. We have no data on the ease of generating sentence completions from the 900+ participants in our cloze norming studies, and if these were available, it would be necessary to identify the underlying factors why one word order would more easily yield verb predictions than another. If it were possible to identify such a factor, however, and demonstrate that at short latencies, the stimuli in Chow et al. (2015) indeed yielded event predictions that only matched one word order, while the stimuli in the present study yielded event predictions compatible with both word orders, the probability-based account could explain why the role-specific verb candidate was predicted in the present study, resulting in a contrast in N400 amplitude. If the search results of the lexical-association mechanisms yield two probable events (e.g. generating both *serve* and *tip*), this reduces the overall certainty

with which either of those events is predicted. However, the generation process based on the full set of available cues (i.e. nouns including argument role information, whether this is represented directly as assigned argument roles or indirectly by way of word order information) yields only one event prediction, which is therefore predicted with higher certainty. Under the assumption that the NNV contexts used in the present study generated competing verb candidates at the lexical association stage of processing, the probability-based account can therefore explain the contrast in N400 outcomes for the present study and past work: in the present study, assuming that the probabilities of the event predictions coming up at the lexical-association stage were evenly matched, the candidate generated by the role-specific mechanism would be exclusively predicted, resulting in an N400 contrast when this candidate was not the sentence target.

The probability-based account is not designed to explain the timing phenomena, and this is a serious shortcoming of the model. In principle, however, it seems reasonable to assume that the predictions generated by a ‘bag-of-arguments’ lexical-associative search mechanism are not generated all at once but instead have different generation timing profiles. It may be the case that the events predicted in the present study’s sentence contexts were generated closer together in time than the events predicted in Chow et al.’s (2015) study, yielding an apparent similar evenness in event distributions in offline data, but exhibiting minute timing contrasts in online prediction. Note that Chow et al. (2018) showed the timing contrast in a specific subset of stimuli whose offline maximum cloze values exceeded a probability of 40% for the verb target. If there is a timing spectrum for the generation of event

predictions, this could encompass super-fast conflicting candidates that trigger the choice of the role-specific candidate at short processing latencies, yielding N400 contrasts, as in the present study; a mix of fast and medium fast candidates, e.g. in the high-predictability items in Chow et al. (2018), such that one candidate is unopposed at short processing latencies (yielding a lack of N400 contrast) but the later emergence of second triggers the selection of a role-specific candidate (resulting in an N400 contrast at longer latencies); and a mix of fast and slow candidates, such as in the low-predictability items in Chow et al. (2018), so that the single candidate generated at the earliest processing stages is never opposed and there is no N400 contrast even at longer processing latencies.

Pending the discovery of a factor that would systematically influence the prediction of verb candidates such that they were generated quickly in the present study but more slowly in Chow et al. (2018), then, the probability-based account could yet provide an explanation for the discrepancy between N400 outcomes in the present study and prior work, as well as explaining the timing contrast found in Chow et al. (2018). However, as the account currently stands, and given the present state of knowledge concerning the contrast in stimulus sets, this is not feasible.

5.6.3 *Future directions*

Ettinger's (2018) calculations of subject-verb context similarity between the present stimuli and the ones used in Chow et al. (2015) suggest that the amplitude of the N400 might be driven by the degree of overlap in the contexts in which embedded

subjects and verbs are found, in that this probability distribution is indirectly or even directly related to a verb candidate's level of activation in prediction. If it is the case that embedded subjects drive verb predictions in NNV object-relative clauses, a high discrepancy between subject-verb cosine relationship should result in an online N400 contrast. At the same time, an ample literature has demonstrated that N400 amplitudes are strongly correlated with offline cloze completion data (Federmeier & Kutas, 1999 and others). The present study's high cloze values reach an average of 36%, whereas the average maximum cloze value in Chow et al. (2015) is 25%. Under an account in which the N400 does not distinguish in a binary fashion between predicted and unpredicted candidates, but reflects nuance in the extent to which an item is predicted, it is possible that the N400 contrast observed in the present study was due to overall higher target cloze values, rather than due to the influence on prediction on some inherent difference in the way that stimuli were created.

We are currently collecting data in a follow-up experiment that directly compares the stimuli from Chow et al. (2015) and a subset of the NNV stimuli used in the present study. The present study's stimuli consists of 60 sets of two low and high cloze pairs. A subset of these stimuli was chosen such that the final stimulus set contains 60 pairs of low and high cloze stimuli, with an average target cloze value of 25% in the high cloze condition and <2% in the low cloze condition, to match the cloze contrast in Chow et al.'s (2015) original stimulus set. However, this subset of stimuli has a higher divergence between the subject-verb cosine relationship than in Chow et al. (2015) or the present study. If this is the factor driving online prediction, there should be a stark N400 contrast between low and high cloze conditions in the

subset of stimuli based on the present study, but not in the stimuli from Chow et al. (2015). By contrast, if the original difference in N400 outcomes was driven by slightly higher target cloze values in the present study, there should be no N400 contrast between low and high cloze conditions in either stimulus set. This investigation will help narrow the scope of possible reasons for the distinction between the results of the present study and prior results from electrophysiological investigations of role-reversed sentences, and provide more evidence to illuminate the nature of predicting events when only event participants are known to the comprehender.

5.7 Conclusion

This investigation tested the predictions of two different prediction-based accounts of the lack of N400 contrast on target verbs in role-reversed NNV and NVN sentence contexts. As predicted by the format mismatch account of verb prediction, we found an N400 contrast between target nouns in canonical and reversed NVN contexts. Unexpectedly, we also found an N400 contrast on target verbs in canonical and reversed NNV contexts. Explorations of the present study's stimuli, and comparison with the stimuli used in Chow et al. (2015), yielded no measurable contrasts between the stimuli in terms of entropy or maximum cloze distribution. However, there was a contrast in offline ratings of the plausibility of canonical and role-reversed sentences. There was also a contrast between the stimuli sets in terms of subject-verb cosine

relationship. Future research will determine to what extent this factor is involved in determining role-specific verb predictions on the basis of argument role information.

6 Argument roles in adult and child comprehension

6.1 Introduction

The experimental investigations in this dissertation differ on a variety of dimensions, and drawing overarching conclusions about sentence comprehension requires a careful investigation of online and offline measures that are thought to index disparate aspects of processing. In one set of studies, we tested children's commitments in online sentence processing, using eye-tracking methodology; in the other, we tested adults' prediction in online sentence processing, using electrophysiological methods. Children's actions probe their comprehension after they have heard the entire sentence, formed a structural parse, interpreted it, and engaged in explicit reasoning to determine the correct response to the experimental task. Studies of prediction in adults, on the other hand, test the extent to which adults are able to use noun and word order to infer argument role information in real time to predict upcoming verbs. The two sides of the experimental investigation in this dissertation share an interest in illuminating how comprehenders are able to use argument role information when it is gleaned from qualitatively different sources: argument role information that includes or excludes verb information, yielding structural parses in the *seal-Agent* or *seal-Eater* formats. The aim of this discussion is therefore to highlight the connection between the two sides of the experimental investigation, and draw some broader conclusions about human sentence processing.

The investigation of German-speaking children's comprehension of passives contrasted the *seal-Agent* commitments that children make on the basis of noun and word order information, exclusive of verbs, with the *seal-Eater* commitments they make when verb information is available. Correctly interpreting a passive sentence involves non-canonically assigning a Patient argument role to the syntactic subject. However, children have a well-documented bias to interpret initial NPs as subjects (see section 2.3), alongside well-documented difficulties with revision (Trueswell et al., 1999 and many others), suggesting that failures in passive comprehension may be related to failures in revising the downstream consequences of an initial subject-first interpretation, which may result in a subject-as-agent parsing commitment (Huang et al., 2017, 2013). Our results revealed a striking variability in children's accuracy in comprehending passives. Passive comprehension outcomes were high (indicating a high rate of accurate argument role revision) when children had committed to a subject-as-agent interpretation, but received a cue to revision prior to the main verb. By contrast, children's passive comprehension outcomes were poor (indicating a lower rate of accurate argument role revision) when the cue to revision was provided after the main verb. Why would a difference in the relative order of the voice cue and the main verb have such a heavy impact on children's ability to revise? I argue in section 6.2 that there is a qualitative difference between argument role information that is or is not tethered to verb information, and lay out why this might make a difference to children's ability to revise.

I pointed out in Chapter 2 that commitment and prediction are largely two sides of the same coin in sentence processing: bottom-up input must be used to build

an interim representation of structure (referred to here as a structural parse), which is then linguistically interpreted and used to access memory and predict upcoming material, including specific lexical items as well as specific linguistic structures. Previous work has not differentiated clearly between linguistic and conceptual sides of language comprehension, despite implicitly appealing to these processes in explaining phenomena in sentence comprehension. In order for a participant to be able to complete typical experimental tasks like answering comprehension questions, the outcome of the linguistic interpretation process must be fed into a process of conceptual interpretation, which interfaces with non-linguistic forms of information such as world knowledge, episodic memory, and allows the comprehender to engage in explicit reasoning on the propositional content of the sentence. Importantly, since parsing and linguistic interpretation are incremental, I posit that conceptual interpretation, too, is incremental, building on the evolving structural parse that is formed from bottom-up input as the sentence unfolds over time. Prediction is dependent on the successes of forming accurate structural parses: information must be integrated in a structural interpretation before it can yield adequate predictions about upcoming sentence material. Conversely, it should be difficult to make predictions on the basis of structural commitments that are in a format which is initially hard to use, because they may be ill-suited to interfacing with the types of non-linguistic information that are required at the conceptual interpretation stage. I will argue that this is the reason for the differences in processing outcomes that this dissertation has uncovered between information in the *seal-Agent* and *seal-Eater* formats. Prior work showing poor online prediction outcomes in role-reversed sentences provides an

example of this. In section 6.3, I explore how the nature of argument role commitments that can be made on the basis of pre-verbal information, and the resulting prediction outcomes, may align with present findings from the realm of child comprehension.

So far, I have argued that qualitative differences in the nature of the argument role information that is extracted from the input (with verb information as in *seal-Eater*, or without as in *seal-Agent*) influence both the structural commitments that comprehenders make on this basis, and the predictions that result from those structural commitments. If it were the case, however, that the lexical categories available to the parser in sentence processing – and specifically, the presence or absence of a verb – determined structural commitment and prediction outcomes, one would expect to see processing outcomes always varying according to the structure of the sentence being processed. However, Chapter 5 shows that this is not the case. My EEG experiment was designed to expand on the hypothesis that the comprehender's ability to use argument role information varies according to the lexical categories on whose basis that argument role information is extracted from the input. The experiment compares N400 amplitudes on the target in canonical and role-reversed NNV and NVN sentence contexts. If it were true that providing verb information made a systematic qualitative change to the nature of argument role information, and that comprehenders are unable to use argument role information without verb information, the expected outcome would be that the N400 would only be sensitive to role information in NVN sentence contexts. While these contexts did demonstrate an N400 contrast, a similar effect was also found in the NNV contexts. This suggests

that there are specific circumstances under which comprehenders are able to use pre-verbal argument role information in the *seal-Agent* format successfully in prediction. In section 6.4, I explore the possibility that arguments are stored with probabilistic information about the frequency distribution of their occurrence under assignment of specific argument roles, and the impact that this would have on processing.

Any model accounting for the data presented in Chapters 4 and 5 would also need to account for the timing factors found in prior work, notably in Chow, Lau, Wang, & Phillips (2018). While timing is not a factor that my experimental manipulations explicitly address, it may have been a confound in the German child comprehension outcomes. In section 6.5 I discuss timing, as distinct from order, as a factor in determining the use of argument role information in comprehension.

Finally, in section 6.6 I formulate a possible framework that could unify the hitherto separate accounts I have laid out for each of the two halves of this investigation – parsing commitments in children, prediction in adults.

6.2 Verb information and argument role commitments

As described in Chapter 4, the fundamental difference between children being able to succeed at revising, or failing in their attempts to revise, a subject-as-agent commitment appears to be whether that subject-as-agent commitment also included verb information or not. I argued in Chapter 2 that comprehenders do have access to argument role information when encountering a subject, even in the absence of a verb: it is not the case that children succeeded at revision in Experiments 4.1 and 4.3

because they made no commitments to a subject-as-agent misinterpretation. Next, I will lay out a possible explanation for why subject-as-agent commitments are easier to revise than commitments to the subject as the agent of a specific, known verb. This explanation draws on the suggestion that conceptual interpretation relies on incremental updates to the comprehender's conscious beliefs about the propositional content of the sentence being processed, and suggests that verbs may play a special role in belief updating in a way that impacts argument role assignment and revision of these assignments.

To what extent can we assume that comprehenders commit to subject-as-agent interpretations? While there is a wealth of evidence that comprehenders commit to parsing initial NPs as subjects (see section 2.3 and references therein), the evidence that interpreting an NP as a subject also means committing to assigning it an agent role is far less certain. Nonetheless, studies of passives have revealed that comprehenders' assumptions about subjects lead into subject-as-agent structural commitments. For these reasons, we assume that comprehenders have access to some form of argument role information prior to encountering a verb: they are able to access a level of granularity which allows them to make the broad distinction between families of roles that are assigned to subjects versus objects, but perhaps they may not initially have access to the finest grain size of argument role information. For instance, the Agent/Experiencer distinction may only be accessible once argument roles have been assigned by the main verb. Nonetheless, the P600 contrast on target verbs in role-reversed sentences (Chow et al., 2015; Kim & Osterhout, 2005) shows that comprehenders have access to some forms of "who does what to whom"

information prior to the verb, even if the precise nuances of the argument roles are not yet available.

We therefore assume that when children participating in Experiment 4.3 and the Subject/Voice condition of Experiment 4.1 encountered the initial NP, they parsed it as a subject, and then formed a further commitment to a subject-as-agent parse. In the input this was immediately followed by the cue to voice. Children's high rate of comprehension in passives indicates that they were successful in revising *seal-Agent* to *seal-Theme*. By contrast, in Experiment 4.2, comprehenders encountered a main verb prior to receiving the cue to voice. Their interpretation therefore includes not just the argument role, but also the verb that this argument role is derived from: that is, upon encountering a verb, comprehenders are committed to an interpretation in which the seal is not just an agent but specifically the agent of eating (*seal-Eater*). The empirical observation is that this form of commitment is more difficult to rescind than a commitment of the form *seal-Agent*, as demonstrated by children's markedly lower accuracy rate in interpreting passives based on structural commitments of the form *seal-Eater*.

There exist a number of possible explanations why this might be the case. One option concerns the concreteness or specificity of the argument role information. Argument role information of the form *Agent* is less concrete than information of the form *Eater*, and it is possible that reanalysing verb-specific argument role information is more difficult. An analogy from the working memory literature may illuminate why. "Chunking" describes a form of concatenating various pieces of information such that they can be stored and manipulated more efficiently as a single unit.

Concatenating different types of linguistic information almost certainly involves building a chunk, and this may bring memory advantages for further processing, as the building-block can be used as a whole, instead of the parser having to use attentional resources in order to unify certain disparate units of structure. However, this memory-saving measure may make a unit more difficult to break apart into constituent pieces for reanalysis, resulting in a lower chance of success in reanalysis and therefore leading to the observed poor comprehension outcomes. According to this explanation, revision happens as the parser builds structural commitments from bottom-up input, but the nature of the information chunks that are being revised partly determines the outcome of revision.

Language comprehension requires the comprehender to build a structural parse and then linguistically and conceptually interpret it. While the previous explanation outlined a way that revision difficulties might arise at the stage of building a structural parse and linguistically interpreting it, it is also possible that the difficulty arises at the conceptual interpretation stage. Studies investigating adults' beliefs about the propositional content of garden-path sentences reveal that misinterpretations linger even after revision has taken place, revealing that the step from parse to interpretation is also susceptible to the effects of difficulty with revision. Christianson et al. (2001) asked comprehension questions after garden-path sentences like "*While the man hunted the deer ran into the woods,*" and found that adults were overwhelmingly likely to incorrectly interpret *the deer* as a direct object complement of *hunted* until well after the end of the sentence. Sturt (2007) found similar effects in garden-path sentences that are purportedly easier to revise

("The explorers found the South Pole was actually impossible to reach"). Christianson et al. (2001) explained their result as a "good enough" parsing effect, suggesting that the interpretation proceeded on the basis of an inaccurate structural parse based on incomplete information. Slattery et al. (2013) expanded this set of findings through a self-paced reading task on garden-path sentences in which a reflexive's gender matched or mismatched an antecedent whose correct identification depended on the resolution of the garden path. They found a gender mismatch effect that indicated that revision had taken place in time for the reflexive to be processed accurately, suggesting that lingering misinterpretation effects are not in fact due to the initial inaccuracies of "good enough" parsing. In a follow-up experiment, a second sentence included a pronoun whose correct resolution depended on correctly revising the garden-path in the first sentence. Slattery et al. (2013) found lingering effects of misinterpretation even on this pronoun, suggesting that although garden-path revision had been completed (as indicated by the gender-mismatch effect), the initial misinterpretation had not been overwritten in memory. Instead, the authors suggested that there is competition between the initial misinterpretation and the interpretation arising from the correct parse, resulting in lingering misinterpretation effects.

None of the above authors formulate an explicit model of the interface between forming a structural parse, linguistically interpreting it, and conceptually interpreting it, with the latter two processes in particular frequently not differentiated (see e.g. Slattery et al., 2013). Experimental results have robustly established that interpretation (linguistic and conceptual) proceeds incrementally, on the basis of structural units of information that are parsed before the entire sentence has been

encountered, given the discrepancy between the flat linear order of linguistic input and the underlying syntactic structure. While bottom-up information is integrated in a structural parse, intermediate structural units are interpreted, leading to belief updates both in terms of the comprehender's unconscious beliefs about the structure of the parse (Kuperberg & Jaeger, 2016) and the comprehender's conscious beliefs about the content of the sentence being parsed. However, once information has passed into the domain of explicit reasoning about the interpretation, the comprehender loses access to information about the precise linguistic source of each element of the belief. For instance, in a sentence like "*While the man hunted the deer ran into the woods,*" comprehenders are not able to trace a belief that <man hunted deer> back to the second NP's structural ambiguity between direct object and embedded subject. If that were the case, competition between beliefs based on the initial misparse and the corrected structural interpretation would be easily resolved, yet Slattery et al.'s (2013) results show that they are not.

This dynamic may also be at play in the present results. In the act-out task employed in Chapter 4, children were required to correctly interpret a sentence as an active or passive, assigning the correct argument role (Agent or Patient) to the subject. They then had to engage in conscious reasoning about the subject and its argument role in order to correctly select the second event participant and act out the event described in the test sentence. (Some children even narrated this process, for instance by explaining to the experimenter that "sharks really like eating seals, I bet that's what happened" before selecting the appropriate toys.) It is possible that the act-out difficulties observed in Experiment 4.2 were due to competition between a

conceptual interpretation based on an initial misparse (following on from an assumption that the subject is the agent), and the conceptual interpretation based on the corrected structural parse (once voice information has been encountered and the structural parse is revised such that the subject is the patient). Children's underdeveloped cognitive control abilities (Mazuka et al., 2009) may be to blame for their difficulty dealing with these competing parses.

A further modification is required to explain why children were able to mediate competing interpretations based on a structural misinterpretation without verb information (*seal-Agent*) but were markedly less successful at doing so when the structural misinterpretation included verb information (*seal-Eater*). It may be the case that the inclusion of verb information makes a structural parse more likely to be conceptually interpreted and passed into the domain of explicit reasoning. One goal of sentence comprehension is to infer the event being described (Kuperberg, 2015), and verbs are event labels. It is therefore possible that an incomplete structural parse is more likely to be interpreted if it includes verb information. Once an incomplete parse is interpreted, it is not replaced by updates; these instead enter into competition with the existing interpretations (Slattery et al., 2013). If this is the case, an incomplete structural parse of the form *seal-Agent* would remain uninterpreted, such that there is no competition with the interpretation of the revised parse. On the other hand, a misinterpretation of the form *seal-Eater* would be immediately interpreted, then re-interpreted after revision to *seal-Eatee* had occurred, resulting in competing interpretations that may be difficult for child comprehenders to mediate between.

6.3 Structural commitments and prediction

The previous section outlined a qualitative contrast in information of the form *seal-Agent* and *seal-Eater*, where the inclusion of verb information in argument role commitments forms a structure whose argument role assignment is harder to revise than when no verb information is included. As suggested earlier, this might be due to effects of chunk size: *seal-Eater* is a larger chunk than *seal-Agent*, and breaking it apart is more difficult, as *seal-Eater* contains two pieces of information (seal is agent; event is eating) rather than one (seal is agent). Alternatively, this might be because verb information makes a structural parse more likely to be interpreted, resulting in a belief update which is difficult to revise because beliefs cannot be easily traced back to the specific units of linguistic information that gave rise to those beliefs. In Chapter 5, I laid out prior results from the role-reversal literature, which showed that adults initially have difficulty using verb-free noun and argument role information in prediction. This section explores a connection between children's ability to revise argument role assignments in the *seal-Agent* format, and adults' apparent inability to use information in this same format to generate verb predictions at short processing latencies.

Initial accounts of the lack of N400 contrast in response to target verbs in role-reversed sentences suggested that parsing follows a dual-route mechanism in which syntactic and semantic information are processed separately, and that in the case of conflict, semantic information would retain the upper hand (Kim & Osterhout, 2005, are best known for espousing this view, but it is prevalent in many other accounts). According to this view, a sentence fragment like "*The rodeo clown knew which bull*

the cowboy had gored” would be processed separately according to its syntax (which indicates that the event being described is <cowboy-Agent gores bull-Patient>) and semantics (which would find the most plausible combination of the arguments and the verb in a <bull-Agent gores cowboy-Patient> event), with the semantic interpretation dominating over the syntactic parse and resulting in a P600 contrast when the parser notes a conflict between the event inferred by the semantic processor and the syntactic processor. The basic view advanced in this family of accounts holds that argument role information is initially ignored in favour of a “good enough” parse that integrates cues from semantic processing, world knowledge, discourse and plausibility.

This explanation has a surprising affinity with previous explanations of children’s difficulty with passives. Many accounts have focused on the idea that children’s linguistic representations are non-adultlike and therefore lead to processing difficulties that result in incorrect comprehension (Borer & Wexler, 1987; Fox & Grodzinsky, 1998), or on the idea that children’s frequency of exposure to passives in the input determines their learning outcomes (Allen & Crago, 1996; Demuth, 1989; Pye & Poz, 1988). The assumption that children do not have access to the syntactic operations that would make it possible for them to correctly assign argument roles in passives is highly similar to “good enough” approaches to online sentence processing, which claim that comprehenders simply ignore some types of syntactic information in the earliest stages of processing. The parallel that scholars typically draw between adult and child processing, summarised as “adults’ first parse is children’s only

parse,”³ is interpreted in representational accounts of children’s sentence processing to mean that young comprehenders appear unable to act on certain types of syntactic information due to deficiencies in their linguistic representation capacities, whereas adults can eventually overcome an initially inaccurate parse. (It must be noted that “good enough” parsing theories are typically sparse on the details of the mechanisms that would later allow adult comprehenders to take into account the types of linguistic information that are initially ignored.)

There is also a class of accounts that surmises that children use plausibility as a guide to argument role assignment. Many studies have found that children have difficulty assigning argument roles in sentences describing “reversible” events (ones where either argument role assignment is equally plausible; Harris, 1976; Maratsos et al., 1985; Turner & Rommetveit, 1967). Strohner & Nelson (1974) argue that children’s sentence comprehension relies on either the “actor-action-object” strategy (Bever, 1970), or the plausibility of argument role assignments in the event being described. Note that this result is not universal: other studies find no effect of reversibility on children’s interpretation of passives (e.g. Aschermann, Gülzow & Wendt, 2004). The argumentation advanced in many of these experimental investigations is that children are ignoring syntactic cues in favour of either a linear-order parsing heuristic or top-down probabilistic inference based on world knowledge. This is essentially the same class of argumentation as the “semantics overrides syntax” account of the lack of N400 contrast in adults’ responses to target verbs in role-reversed sentences. In both families of argumentation, the assumption is

³ Attributed to John Trueswell.

that comprehenders do not have access to argument role information from the bottom-up input; in accounts of children's difficulty with passives, this may be reduced to deficits in linguistic representation, whereas in accounts of adults' difficulty with role-reversed sentences, this is due to "good enough" parsing which initially ignores word order cues to argument role assignment. In the child comprehension literature, one reason to regard this account with suspicion is the very variability of the results: if children always prioritised plausibility cues over syntactic input, this result should display less variation in experimental outcomes than is actually found across studies. This variability in experimental outcomes is plausibly introduced by the difficulty of clearly portraying argument roles in reversible actions like "hugging" (what is the visual difference between a picture where "Tom hugs Mother" and "Tom is hugged by Mother"?), a methodological issue which undermines the comparability of results from a literature that has been largely dominated by picture-choice tasks (see Chapter 4 for further discussion). Nonetheless, a range of authors employ what is essentially a "good enough" explanation of children's difficulty with passives: early in development, young comprehenders ignore cues to argument role assignment and instead default to other, sometimes extralinguistic, sources of information to interpret complex sentences.

While both adults and children are prone to interpret initial NPs as subjects, there is doubt over the extent to which this leads them to commit to a subject-as-agent interpretation. For the purposes of this discussion I assume that children, like adults, can and do extract information about argument roles from cues like word order, and that they make parsing commitments on this basis. As outlined in Chapter 2, there is a

clear link between making a structural commitment and generating predictions. In order to use sentence material and generate viable predictions for upcoming words, the comprehender must build a structural parse of the material that has already been encountered. Chapter 4 discussed evidence that children are able to revise argument role assignments in the *seal-Agent* format, but do so less successfully in the *seal-Eater* format. Prior results from the role-reversal literature (Chow et al., 2015; Hoeks et al., 2004; van Herten, Chwilla, & Kolk, 2006; van Herten, Kolk, & Chwilla, 2005 and many others) suggest that adults have difficulty generating role-specific verb predictions from precisely this information format. The same information format, *seal-Agent*, is both easy to revise and difficult to use in prediction. It is possible that the factors preventing *seal-Agent* from being conceptually interpreted and resulting in an update to the comprehender's explicit beliefs about the proposition are related to the factors preventing *seal-Agent* from being used effectively to generate verb predictions. The account laid out in Chow et al. (2016) suggests that events in memory are labelled by verbs and their likely participants, but do not provide easy access to relational information such as the argument roles that the participants take, resulting in difficulty finding a suitable verb candidate. If events in memory cannot be accessed through only a combination of noun and argument role information, it may be the case that conceptual interpretation cannot proceed either.

According to this combined account, when encountering a noun, both children and adults use word order information as a cue to infer argument role assignment, resulting in structural commitments in the *seal-Agent* format. However, processing further downstream, such as prediction, conceptual interpretation and updates to the

comprehender's explicit beliefs about the proposition, all depend crucially on verb information. In the case of children's comprehension of passives, this turns out to be a blessing in disguise: the delay in conceptual interpretation and belief updating provides a narrow window of time in which the cue to voice can be integrated such that argument roles are correctly reassigned, the corrected parse can be interpreted, and the child acts out the correct role assignment (Experiments 4.1 and 4.3). In the case of adults' predictive processes in role-reversed NNV object-relative clauses, this is precisely the hang-up: because *seal-Agent* does not include the verb information that would make it possible to use argument roles in prediction, the resulting verb candidates are not specific to the argument roles of the preceding NPs.

Finding a common backbone across adult and child comprehension of argument roles is a highly attractive outcome. Yet if it were the case that the utility of argument role information in prediction were fully determined by the inclusion or exclusion of verb information, we would expect prediction to always fail when verb information is excluded. However, as the results of Experiment 5.1 demonstrate, this is not the case. The following section explores a set of probabilistic factors that might explain the outcome of Experiment 5.1.

6.4 Probabilistic factors in commitment and prediction

The EEG experiment described in Chapter 5 was designed to replicate and extend the results of a body of prior literature on verb prediction in role-reversed sentences. The specific aim was to test the hypothesis that verb information makes it possible to

generate role-specific candidates for event participants, examining whether argument role information in the *seal-Eater* format could overcome the format mismatch that is hypothesised to underlie the initially role-insensitive N400 seen in, among many others, Chow et al. (2018, 2015). In the process, we discovered that under specific circumstances, argument role information in the *seal-Agent* format can in fact be used at short latencies to predict role-specific verb candidates. As rigorous and extensive post-hoc analyses showed (Chapter 5), the precise nature of these circumstances has proved difficult to determine. Here, I explore the possibility that probabilistic factors may tilt the balance in certain lexical items, resulting in a change in processing outcomes that yields the observed N400 contrast.

A relevant parallel from the garden-path literature involves verbs that have the capacity to take both direct object and sentential complements, but individual verbs vary in terms of which complement they typically take. A variety of studies has shown that adults' processing follows these probability distributions: direct object bias verbs incur garden-path effects when disambiguation reveals the ambiguous NP to be the subject of a sentential complement, and vice versa (Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Trueswell, Tanenhaus, & Kello, 1993, and others). Discussions of these phenomena have tended to focus on the question of whether frequency-based information influences how comprehenders parse bottom-up input (constraint-based models; see e.g. Trueswell et al., 1994), and it seems fairly straightforward that this is indeed the case. In order for processing to proceed in this manner, it must be the case that lexical entries for verbs are stored with information as to the probability distributions of different types of complement. When a

comprehender encounters a verb like “believe” prior to an NP, the processor must entertain the possibility of two possible structural parses of that NP. If it were the case that both parses were entertained as equally likely, purely on the basis that both are possible, there should be no garden-path contrasts at all when the comprehender encounters disambiguating information: it should be equally easy or difficult to integrate this disambiguating information in either possible parse, and discard the irrelevant one. Likewise, if there existed a global preference for one parse over the other, we would expect to see unidirectional garden-path effects: for instance, if the bias were a uniform preference for sentential complements, comprehenders should always experience processing difficulty in disambiguating regions that indicated that the ambiguous NP is a direct object. This again is not the case.

The results from Chapter 5 bear some similarity to these findings. In this experiment, comprehenders showed an ability to use argument role information in order to predict a role-specific verb. Lexical entries for verbs that can take different complement types are stored with probabilistic information about the likelihood of different structural parses, and this probabilistic information can tip the balance in terms of how the processor integrates upcoming sentence material. It is possible that a similar principle is at work with nouns, which may be stored with probabilistic information about which types of argument role they most frequently take. This is compatible with observations from studies investigating the manner in which comprehenders predict upcoming event participants. McRae, Amyote & Ferretti (1997) found distinctive agenthood and patienthood preference profiles for various nouns assessed through questions like, “How common is it for a monster to

frighten/be frightened by something?” A modified self-paced reading task revealed that adjective modifiers can push comprehenders’ perception of what nouns and verbs combine to make the noun a good agent of the verb. In sentences like “*The shrewd heartless/young naïve gambler was manipulated by the dealer*”, the authors found elevated reading times on “dealer” when the adjectives suggested that the noun was a poor agent of *manipulate*. While these examples show how agenthood and patienthood preference can be impacted by combination with a verb, it seems plausible that this information might in part be stored on the noun as well. There are other featural sources of information, such as animacy, that influence whether a noun is perceived as an agent or patient. For instance, MacWhinney, Bates, & Kliegl (1984) found that animacy was a determining factor in adult comprehenders’ decision to identify nouns as agents in English, German and Italian. In comprehending object-initial sentence orders, German-speaking children are better able to use case marking cues when they align with animacy (Dittmar et al., 2008), suggesting that children have a preference for aligning subjects with agenthood and animacy. In adult online sentence processing, the evidence is not always clear-cut, but it seems that inanimate agents result in distinct ERP profiles (Philipp, Bornkessel-Schlesewsky, Bisang, & Schlewsky, 2008). Trueswell, Tanenhaus & Garnsey (1994) found that typicality ratings for the agenthood or patienthood of initial NPs were correlated with reading times on disambiguating regions following an ambiguous verb (*The defendant/evidence examined by the lawyer...*), indicating that animate NPs led to slower reading times in the disambiguating regions. This suggests that animacy and subjecthood combine into a subject-as-agent commitment which leads to processing

difficulties when the disambiguating region reveals that the initial NP is the subject of a passive, and therefore a patient.

If it is the case that nouns are stored with probability distributions over their likelihood of taking an agent or patient role, this might impact how verb candidates are generated. For instance, if *bull* and *cowboy* are equally likely to be agents, it may be easier to generate verb candidates that are specific to the role assignment in question, if this generates strong competing verb candidates. By contrast, nouns like *villager* and *ghost* may have more divergent probabilities of being an agent, resulting in only one strong role-specific verb prediction. This might be because these probability distributions over argument roles promote greater certainty for the argument role commitments in the structural parse. This may be especially relevant in the case of English NNV clauses, which require revision, due to the bias to interpret the initial NP of an embedded clause as a subject. This in turn could help generate role-specific verb predictions more quickly by giving an activation boost to role-specific verb candidates, and possibly giving a time lead to the role-specific serial search mechanism, resulting in a shorter latency for generating role-specific verb candidates.

Future research will have to determine to what extent nouns are indeed stored with probabilistic information concerning argument roles. There is some tentative support for this possibilities in the findings of Ettinger (2018), who compared the vector space similarity of embedded subject nouns and verbs (subject-verb cosine relationship) in NNV sentence frames used in Chow et al. (2015) and in the EEG experiment I describe in Chapter 5. Her results suggest that the embedded subjects in

Chow et al. (2015) were found in similar contexts as target verbs across both the canonical and reversed conditions, whereas there was a greater divergence for the same measure in the present study's NNV stimuli. For Chow et al.'s (2015) stimuli, this means that the embedded subject of a canonical sentence (e.g. *ghost*) was just as likely to co-occur with *haunt* as the embedded subject of a reversed sentence (e.g. *villager*). By contrast, in the stimuli used in the present work, *bull* was more likely than *cowboy* to co-occur with *gore*.

Vector space similarity is calculated in terms of the co-occurrence of the two words being compared, and this does not include any information about the structural relationship between these two words in the contexts that are being used as data points for calculating vector space similarity. However, it seems plausible to consider a noun's subject-verb cosine relationship a reasonable proxy for its suitability as an agent of that verb. Therefore, a similarity between these values across high and low cloze conditions (as in Chow et al., 2015) indicates that both nouns are similarly compatible as agents of that verb, whereas a discrepancy (as in the present study) indicates that one noun is markedly more suited to acting as an agent of that verb than the other.

One possibility is that nouns are stored with a distribution of probabilities over their likelihood of being the agent or patient of a specific verb. For example, *bull* and *cowboy* type nouns from the present experiment's stimuli might be stored with probabilities that look somewhat like the ones in Table 6.1.

Table 6.1: Mock distribution of nouns over argument roles with specific verbs (Ehrenhofer stimuli)

	<i>-as-Agent</i>	
Bull	Gorer	30%
	Trampler	20%
	Kicker	10%
	Lassoer	0%
	Whipper	0%
	Rider	0%

Cowboy	Rider	35%
	Lassoer	20%
	Whipper	20%
	Trampler	0%
	Kicker	10%
	Gorer	0%

On the other hand, *villager* and *ghost* type nouns might be stored with probabilities similar to the ones in **Error! Unknown switch argument.**

Table 6.2: Mock distribution of nouns over argument roles with specific verbs (Chow et al., 2015 stimuli)

	<i>-as-Agent</i>	
Ghost	Haunter	30%
	Scarer	20%
	Seer	20%
	Exorciser	0%

Villager	Seer	35%
	Exorciser	10%
	Haunter	10%
	Scarer	20%

The chief contrast between the two sets of probability distributions is that for the Ehrenhofer stimuli, there is a clear contrast between the identity of the verbs that are good fits for the agenthood of either noun, whereas for the Chow stimuli, there is greater overlap. In generating verb predictions, this could mean that the initial lexical-associative mechanism generates all verb candidates through this pattern of distributions. In the Chow et al. (2015) stimuli, this results in the generation of a

similar group of verbs, resulting in facilitated N400s for these targets even when they do not match the role assignment as well as other candidates. These can only be differentiated at later processing latencies through the activity of the role-specific serial search mechanism, which checks whether the argument role assignments of both nouns fits a given verb candidate.

In the Experiment 5.1 stimuli, on the other hand, the distribution over probabilities of argument roles with specific verbs generates different pools of verb candidates in the lexical-associative generation mechanism, as there is less overlap in terms of the agent-of-verb candidates stored with the two nouns. This means that role-specific verbs (e.g. in the case of *bull*, this would be *gore*, from the high probability of *Gorer*) will have higher levels of activation than non-role-specific verbs (e.g. *ride*, from the low probability of *Rider*). This would result in a contrast in facilitation for *ride* vs. *gore* in the *which cowboy the bull had...* prediction context. The later stage of checking through the role-specific search mechanism then serves to confirm, rather than rectify, the verb prediction.

This tweak to the Chow et al. (2016) account preserves some of the account's crucial features, while ameliorating those aspects of the account that were simply unable to explain the results from Chapter 5. The important insight from that account is that argument role information may be difficult to use in prediction if it is divorced from verb information. Note that in my account, it is not the case that nouns are stored with a distribution over argument roles (e.g. *bull* [80% Agent, 20% Patient], etc.). If that were the case, we would expect these assignments to always dominate verb generation, such that no matter what the actual argument role

assignment, the verbs generated at the lexical-associative stage are mostly compatible with *bull-Agent* (even in the context of *which bull the cowboy had...*). If this were the case, we would expect no N400 contrast between the canonical and reversed conditions, as the same verb would be generated in either order. However, by positing that nouns are stored with combined verb-argument roles (*Gorer, Rider, Hauntee, etc.*) we provide the possibility of a more nuanced distribution over likely upcoming verbs, which is able to take the current argument role assignment into account from the bottom-up input, but is also sensitive to probabilistic variability in the verbs that these nouns co-occur with.

The idea that lemmas are stored with fine-grained information about the contexts they appear in is not unprecedented. As discussed in section 2.3, verbs are stored with probabilistic distributions over complements (e.g. direct objects or sentential complements, Trueswell et al., 1994). Similarly, homonyms are stored with the probabilities of their different meanings (Duffy, Morris, & Rayner, 1988; Sereno, Pacht, & Rayner, 1991). Probability distributions over verb-combined argument roles may be able to address the dilemma from Chapter 5, which is that accounting for the results of Chapter 5 alongside a long line of role reversal studies that have shown the N400 to be insensitive to argument role information requires some mechanism that can differentiate between cases where argument role information is ignored, and cases where it is taken into account in verb prediction. Simple probabilities of nouns being likely participants of specific events, as suggested by Kuperberg (2016), does not allow for the timing effects found in Chow et al. (2018), nor does it provide a systematic reason for the contrast between the NNV results from Chow et al. (2015)

and the present EEG study. However, a combination of the two accounts, resting on observations from the formation of structural commitments and the influence of probabilistic information on parsing decisions, suggests that the noun pairs in the present study may have been more balanced in terms of their probability of being agents.

One important caveat to consider in further developing this account is the results of the substitution conditions in Chow et al. (2015). Chow et al.'s (2015) low-cloze conditions came from two different sources. In the reversal manipulation, target verbs had a cloze probability of 0% due to the reversal of the preceding nouns (... # *which waitress the customer had served*). In the substitution manipulation, target verbs had a low cloze probability because of the replacement of the first NP with one that could not be a participant in the event denoted by the verb (... *which tenant/ # realtor the landlord had evicted*). In the substitution manipulation, the identity of the subject NP was kept constant across both conditions, yet there was an N400 contrast between these two conditions. If subject-verb cosine relationship were the primary driver of N400 amplitude, this would predict a lack of contrast when the subject is held constant between conditions. However, the results from Chow et al. (2015) may complement, rather than undermine, the subject-verb cosine relationship account of the N400 contrast in Experiment 5.1. Under the account of verb prediction advanced in Chow et al. (2015, 2016), an initial stage lexically pre-activates all verbs that are compatible with both NPs in the clause. In the case of the substitution conditions, the two conditions use different lexical items (tenant + landlord vs. realtor + landlord). While the subject itself (landlord) is constant across these two conditions, the identity

of the participants available for prediction leads to the generation of disparate verb predictions, and this part of the verb generation process may be the source of the N400 contrast. On the other hand, in the reversal conditions in Chow et al. (2015), as well as the NNV conditions of Experiment 5.1, the same noun phrases are used, but in a different word order. In these cases, the difference in subject-verb cosine relationships among conditions is a further contributor to the amplitude of the N400, such that only a contrast in subject-verb cosine relationships between conditions (as in Experiment 5.1) leads to a difference in N400 amplitudes. While there were N400 contrasts on target verbs in both the Chow et al. (2015) substitution conditions and the NNV reversal conditions in Experiment 5.1, they could plausibly be derived from different sources: the disparate identities of the pair of NPs being used to generate verb predictions in the Chow et al. (2015) substitution conditions, and the disparate subject-verb cosine relationship in the reversal conditions in Experiment 5.1.

Data collection is currently underway to replicate the present findings and the results of Chow et al. (2015) in the same experiment. This experiment uses the NNV role-reversal stimuli from Chow et al. (2015) and a subset of stimuli that had attempted to replicate the same condition in the present study. In a key departure from the present experiment, the subset stimuli's canonical condition has an average high cloze probability of 25% to match the cloze probabilities in Chow et al. (2015). However, the subset of stimuli taken from Experiment 5.1 have a greater divergence in the subject-verb cosine relationship between the canonical and reversed conditions. If subject-verb cosine relationship does indeed predict the amplitude of the N400

response, this would further support the view that probability distributions over combined verb-argument role information influence prediction.

6.5 Timing in commitment and prediction

Timing is a final factor that is likely to impact the outcomes of building structural commitments and prediction. As briefly discussed in Chapter 4, the relative order of the voice cue and the main verb were confounded with the duration of intervening words. That is, in Experiments 4.1 and 4.3, children heard the cue to voice immediately prior to or following the subject; any subject-as-agent interpretation would therefore only exist unchallenged until the appearance of the voice cue in the input (in Experiment 4.1, this would be a duration of roughly 540 ms across all conditions). By contrast, in Experiment 4.2, the voice cue was the last word in the sentence; any subject-as-agent interpretation therefore existed for the majority of the sentence (on average, across all items and conditions, this was a duration of 3270 ms). However, up until the final word, the comprehender receives more information that is compatible with a subject-as-agent misinterpretation (Barbara Höhle, p.c.). Under the assumption that revision becomes more difficult when the weight of evidence is compatible with the original parse, and there is no greater weight of evidence in favour of the revised parse, it may be the case that the post-verbal revision difficulty we observed in Chapter 4 is due to the lapse of time, rather than the nature of the information occurring in the interim. Alternatively, the simple passage

of time itself may cement a parse in memory, making it more difficult to revise when a cue to reanalysis is encountered.

Evidence from the study of lingering misinterpretations corroborates this effect, although it is impossible to dissociate the effects of time and the accumulation of consistent evidence in these studies. Christianson et al. (2001) compared comprehension outcomes for garden-path sentences in which the ambiguous head noun was either followed or preceded by modifying information (“*While Harry chewed the steak that was brown and juicy fell to the floor*”/“*While Harry chewed the brown and juicy steak fell...*”) and found that adult comprehenders were reliably more prone to lingering effects of misinterpreting the ambiguous noun as a direct object when the head preceded the modifying information. It is possible that either the quantity of compatible evidence, or the simple passage of time, make an incomplete parse more likely to be interpreted and passed into the domain of explicit reasoning. The reason for revision difficulty could then still be essentially the same as the suggestion I laid out earlier: conceptual interpretations are not labelled with their exact linguistic source, and revising a structural parse would therefore not necessarily be straightforward to reconcile with revising a conceptual interpretation, resulting in overall misinterpretation.

In the EEG literature, too, timing is an important factor. Chow et al. (2018) show that an N400 contrast arises in role-reversed NNV sentences in Mandarin if there is a delay between the nouns and the final verb. Momma (2016) showed that a temporal delay introduced through a longer SOA between the role-reversed noun and verb resulted in an N400 contrast, though no such contrast appeared at short SOAs.

The suggestion advanced in Momma (2016) and Chow et al. (2016) is that this delayed contrast highlights a delay in the parser's use of argument role information in prediction, resulting from a mismatch between the noun-argument role (*seal-Agent*) search probe format and the way that events are stored in memory.

Note that the timing contrast in Chow et al. (2018) was only found in a subset of experimental items with <40% cloze probabilities on the target verb. If the process of generating role-specific predictions takes time due to a format mismatch, it is possible that some combinations of nouns are not strongly predictive enough to generate a strong role-specific verb candidate even in offline measures. Conversely, if specific combinations of nouns are highly predictive of strong verb candidates (for the probabilistic reasons outlined above), it may be the case that this causes predictions to be generated at shorter latencies. Timing was not explicitly manipulated in the EEG experiment I outlined in Chapter 5, but there was a difference in high cloze values (35% probability in high cloze conditions, compared to 25% in Chow et al., 2015, and Chow et al., 2018). In addition, the contrast in stimulus creation led to polarised offline verb predictions: each noun order yielded offline verb predictions that were not generated by the opposite order. Although the effects of this difference in stimulus creation have eluded quantification, whatever the underlying factor, it led to verb predictions being generated even at the short processing latencies tested in my EEG experiment. It is possible that these stimulus differences led to a distinct temporal profile in prediction, such that role-specific verb candidates were generated at an earlier point in the prediction process than in Chow et al. (2015), Chow et al. (2018) and Momma (2016).

If this is the case, the results of the present EEG study are explained as a combination of probabilistic factors and the format mismatch account. If nouns are stored with probability distributions over argument roles, more polarised probability distributions may lead to more rapid online predictions than less polarised probability distributions. However, it should still be the case that prediction follows a temporal profile in which the earliest stages are role-insensitive, building towards greater role-sensitivity as time progresses due to the operation of the role-specific serial search mechanism. If so, the stimuli used in the EEG experiment in Chapter 5 should yield a lack of N400 contrast at shorter latencies, e.g. if the relative clauses were presented in the simple past instead of past perfect (*which bull the cowboy gored* instead of ... *had gored*).

6.6 Parallels in adult and child argument role processing

So far, I have attempted to lay out theories of argument role processing that stay largely within the confines of each of the experimental investigations in this dissertation, and their individual larger context. The purpose of this section is to briefly sketch out a speculative unifying framework of argument role processing which is able to account for the phenomena that we have so far been discussing individually in the context of children's processing of argument roles in passives and adults' processing of argument roles in role-reversed embedded relative clauses.

In the child studies, we saw that when the intermediate parse includes a noun and its argument role information only (*seal-Agent*), children are able to successfully

revise that argument role assignment, whereas when the intermediate parse also includes verb information (*seal-Eater*), argument roles are less likely to be successfully revised. In adults, we saw that when the intermediate parse includes a noun and its argument role information, comprehenders' ability to generate role-specific verb predictions depends on that noun's stored distributions over combined verb-argument role information (*bull-Gorer*, *bull-Rider*, etc.). The two sets of phenomena share a key sticking-point, which is that bottom-up information is accessed in the *NP-Agent* form, but further processing (measured by children's interpretations and adults' predictions) depends on information that is stored in the *NP-Verber* form.

In the case of children, we saw that subject-as-agent misinterpretations selectively lingered only when verb information had been encountered. We hypothesised that conceptually interpreting a parse requires a format that can bridge the divide between linguistic information (such as argument roles) and world knowledge (such as events), and that verb information provides the missing link between these two types of information by making it possible to combine information into the *seal-Eater* format. In adults, we suggested that this linked information format is involved in predicting upcoming events. In children, we saw that as long as this linked information format had not yet been created, comprehenders readily revised argument role assignments.

Essentially, the framework that I suggest here is one in which linguistic information must be translated into a different format in order to be used in interpretation or prediction, both of which require access to world knowledge.

However, this format has the added characteristic of being difficult to break back down into its individual components for repair when revision is required. We see the outcomes of this linking format as an advantage when children are able to revise complex argument role assignments before verb and argument information have been combined. But it is also a disadvantage, such as when the absence of verb information leaves adult comprehenders at the mercy of probability distributions over elements in this format, oftentimes generating inaccurate verb candidates in prediction.

6.7 Conclusion

In this chapter, I explored the possibility of a qualitative distinction between argument role information that is or is not combined with verb information (*seal-Agent* vs. *seal-Eater*), and laid out an account of how this contrast might affect processing in adults and children. I suggested that the combination of verb and argument role information in the *seal-Eater* format might be harmful children's processing of passives, in that it prevents them from being able to revise argument role assignment when they encounter a cue to voice. In adults, by contrast, this format is required in order to be able to generate role-specific verb predictions, meaning that comprehenders who are generating expectations of upcoming verbs must instead rely on probability distributions over past instances of the noun in such constellations. This can lead to prediction failure when nouns' probability distributions over verb-argument combinations do not sufficiently distinguish between the noun as agent or patient of a particular verb (Chow et al., 2015; Ettinger, 2018), or, more rarely,

prediction success when this probabilistic information does align to allow comprehenders to generate role-specific verb predictions.

7 Conclusion

In online comprehension, extracting information about who did what to whom is crucial to understanding sentences in real time. This task is challenging for both child and adult comprehenders. The aim of the experimental investigations in this dissertation was to shine a light on the processing of two constructions which are known challenges in comprehension: passives, in which non-canonical argument role assignment results in a subject being assigned a patient role instead of an agent, and which is difficult for child comprehenders to process; and role-reversed embedded NNV clauses, in which adult comprehenders have difficulty quickly using argument role information to predict upcoming sentence material.

My investigation of German five-year-olds' processing of passives demonstrated a surprising variability in their comprehension outcomes. Children's behavioural responses showed a high level of accuracy in comprehending passives when the cue to voice (the auxiliary *wurde*) was presented prior to the main verb, but they did markedly worse when the cue to voice was presented after the main verb. This is a departure from previous results in a number of ways. It goes against the established wisdom that children generally fare poorly in comprehending passives (due to a combination of passives' low frequency, or their deficient linguistic representations), and instead pins the blame on the difficulty of revising argument role assignments, which I show to be affected by the provision or absence of verb information. The result also provides much needed clarification for prior work suggesting that German children may perform better in passive comprehension (see Aschermann et al., 2004 for an example) for a variety of reasons including flexible

word order. This is partially true – as I show, the V2 word order in which the auxiliary precedes the main verb does indeed lead to better comprehension outcomes – but I also show that German verb-final word order brings about the same comprehension disadvantages faced by children learning other languages. This set of studies contributes to the child language processing literature in demonstrating that children’s ability to process complex constructions is graded, and that even “ballistic parsing” can take different forms depending on the type of information that comprehenders are able to use in processing.

For adult comprehenders, too, my investigations pinpointed variability in the use of argument roles in online processing. Prior work (see Chapter 5 and references therein) demonstrated that adults struggle to use noun phrases and their argument roles to predict target verbs, possibly due to a format mismatch between search probes in the *NP-Role* format and the format of events in memory (Chow et al., 2016). I found that under specific circumstances, adult comprehenders are in fact able to use argument role information in prediction: due to certain probabilistic factors, or through the inclusion of verb information in the context. This study contributes to the adult processing literature in demonstrating that adults’ difficulty in processing argument role reversals stems from a fragility in using this information immediately, rather than an inability to do so.

We saw in these findings the beginnings of an account of argument role processing where comprehenders extract some information about argument roles from incoming sentences and immediately build structural parses which accurately reflect these commitments. However, conceptually interpreting this parse, as well as

generating a prediction from it, requires a translation into a format that can interact with both linguistic memory and world knowledge, and verbs may provide the bridging point between these two different forms of memory. The child studies show data that are consistent with a view in which pre-verbal argument role commitments can be revised because they have not yet been conceptually interpreted, whereas post-verbal argument role commitments pose more difficulty for revision because they have been conceptually interpreted and therefore compete with each reinterpretation on the basis of a revised parse. In adults, prior work shows data that are consistent with a lack of access to event memory, resulting in non-role-specific predictions of verb targets from argument roles.

It also appears that at some level of specificity, associative probabilities between different concepts have the power to overcome difficulties with prediction that arise as a result of format mismatch. We saw this in the adult comprehension study, where embedded subjects' subject-verb cosine relationship may have been related to participants' ability to predict target verbs on the basis of preceding nouns in NNV clauses. While the exact mechanisms here are not yet understood, further research is underway to examine whether this is indeed a contributing factor in prediction.

My work shows that both child and adult comprehenders are more skilled at using argument role information than they have previously received credit for. On a broader level, I have drawn a parallel between adults' and children's processing, and between committing to an argument role structure and using it in prediction. Both adults' and children's ability to use argument role information in real-time sentence

comprehension is fragile. In children, we saw this fragility as a positive: in the absence of main verb information, children were able to revise initial subject-as-agent misinterpretations, leading to improved comprehension outcomes in passives. In adults, past results suggested that the precise moment in which it is possible for children to revise argument role information with relative ease is also the moment in which adults have difficulty using this information in prediction. However, here, too, we saw variability: when argument role information is combined with verb information, or when probabilistic factors align with argument role information to generate strong verb predictions, adult comprehenders are able to use argument role information in prediction. These findings once again highlight the fundamental similarity in children's and adults' language processing.

Appendix 1: Critical trial materials for Experiment 4.1

Conditions:

- A pre-subject active
- B pre-subject passive
- C post-subject active
- D post-subject passive

Item set	Condition	German	English
1	A	Heute hat die Robbe ihn gefressen.	Today the seal ate it. (= fish)
	B	Heute wurde die Robbe von ihm gefressen.	Today the seal was eaten by it. (= shark)
	C	Die Robbe hat ihn heute gefressen.	Today the seal ate it. (= fish=)
	D	Die Robbe wurde heute von ihm gefressen.	Today the seal was eaten by it. (= shark)
2	A	Gestern hat der Hund ihn verfolgt.	Yesterday the dog followed it. (= hare)
	B	Gestern wurde der Hund von ihm verfolgt.	Yesterday the dog was followed by him. (= hunter)
	C	Der Hund hat ihn gestern verfolgt.	Yesterday the dog followed it. (= hare)
	D	Der Hund wurde gestern von ihm verfolgt.	Yesterday the dog was followed by him. (= hunter)
3	A	Gerade hat der Frosch ihn gefangen.	Just now the frog ate it. (= beetle)
	B	Gerade wurde der Frosch von ihm gefangen.	Just now the frog was eaten by it. (= stork)
	C	Der Frosch hat ihn Gerade gefangen.	Just now the frog ate it. (= beetle)
	D	Der Frosch wurde Gerade von ihm gefangen.	Just now the frog was eaten by it. (= stork)
4	A	Eben hat der Junge ihn getragen.	Just now the boy carried it. (= bunny)
	B	Eben wurde der Junge von ihm getragen.	Just now the boy was carried by him. (= father)
	C	Der Junge hat ihn eben getragen.	Just now the boy carried it. (= bunny)
	D	Der Junge wurde eben von ihm getragen.	Just now the boy was carried by him. (= father)
5	A	Heute hat das Huhn ihn gefressen.	Today the chicken ate it. (= worm)
	B	Heute wurde das Huhn von ihm gefressen.	Today the chicken was eaten by it. (= fox)
	C	Das Huhn hat ihn heute gefressen.	Today the chicken ate it. (= worm)

	D	Das Huhn wurde heute von ihm gefressen.	Today the chicken was eaten by it. (= fox)
6	A	Gestern hat die Katze ihn erschreckt.	Yesterday the cat scared it. (= bird)
	B	Gestern wurde die Katze von ihm erschreckt.	Yesterday the cat was scared by it. (= dog)
	C	Die Katze hat ihn gestern erschreckt.	Yesterday the cat scared it. (= bird)
	D	Die Katze wurde gestern von ihm erschreckt.	Yesterday the cat was scared by it. (= dog)
7	A	Gerade hat der Junge ihn zertrampelt.	Just now the boy trampled it. (= beetle)
	B	Gerade wurde der Junge von ihm zertrampelt.	Just now the boy was trampled by it. (= elephant)
	C	Der Junge hat ihn gerade zertrampelt.	Just now the boy trampled it. (= beetle)
	D	Der Junge wurde gerade von ihm zertrampelt.	Just now the boy was trampled by it. (= elephant)
8	A	Eben hat der Feuerwehrmann ihn gerettet.	Just now the fireman rescued him. (= boy)
	B	Eben wurde der Feuerwehrmann von ihm gerettet.	Just now the fireman was rescued by it. (= helicopter)
	C	Der Feuerwehrmann hat ihn eben gerettet.	Just now the fireman rescued him. (= boy)
	D	Der Feuerwehrmann wurde eben von ihm gerettet.	Just now the fireman was rescued by it. (= helicopter)
9	A	Heute hat der Dieb ihn entdeckt.	Just now the thief discovered it. (= treasure chest)
	B	Heute wurde der Dieb von ihm entdeckt.	Just now the thief was discovered by him. (= policeman)
	C	Der Dieb hat ihn heute entdeckt.	Just now the thief discovered it. (= treasure chest)
	D	Der Dieb wurde heute von ihm entdeckt.	Just now the thief was discovered by him. (= policeman)
10	A	Gestern hat die Maus ihn gefressen.	Yesterday the mouse ate it. (= cheese)
	B	Gestern wurde die Maus von ihm gefressen.	Yesterday the mouse was eaten by it. (= owl)
	C	Die Maus hat ihn gestern gefressen.	Yesterday the mouse ate it. (= cheese)
	D	Die Maus wurde gestern von ihm gefressen.	Yesterday the mouse was eaten by it. (= owl)
11	A	Gerade hat der Stein ihn zerschlagen.	Just now the rock smashed it. (= jug)
	B	Gerade wurde der Stein von ihm zerschlagen.	Just now the rock was smashed by it. (= hammer)
	C	Der Stein hat ihn gerade zerschlagen.	Just now the rock smashed it. (= jug)
	D	Der Stein wurde gerade von ihm zerschlagen.	Just now the rock was smashed by it. (= hammer)
12	A	Eben hat der Junge ihn abgeschleckt.	Just now the boy licked it. (= lollipop)
	B	Eben wurde der Junge von ihm abgeschleckt.	Just now the boy was licked by it. (= dog)
	C	Der Junge hat ihn eben abgeschleckt.	Just now the boy licked it. (= lollipop)
	D	Der Junge wurde eben von ihm abgeschleckt.	Just now the boy was licked by it. (= dog)

Appendix 2: Critical trial materials for Experiment 4.2

Conditions:

A post-verb active

B post-verb passive

Target objects are as in Experiment 4.1.

Item set	Condition	German	English
1	A	Zeig mal, wie die Robbe gebissen und gefressen hat.	<i>Show me how the seal bit and ate.</i>
1	B	Zeig mal, wie die Robbe gebissen und gefressen wurde.	<i>Show me how the seal was bitten and eaten.</i>
2	A	Zeig mal, wie der Hund gesucht und gejagt hat.	<i>Show me how the dog searched and chased.</i>
2	B	Zeig mal, wie der Hund gesucht und gejagt wurde.	<i>Show me how the dog was searched for and chased.</i>
3	A	Zeig mal, wie der Frosch geschnappt und gefressen hat.	<i>Show me how the frog snapped and ate.</i>
3	B	Zeig mal, wie der Frosch geschnappt und gefressen wurde.	<i>Show me how the frog was snapped up and eaten.</i>
4	A	Zeig mal, wie der Junge gesehen und gestreichelt hat.	<i>Show me how the boy saw and cuddled.</i>
4	B	Zeig mal, wie der Junge gesehen und gestreichelt wurde.	<i>Show me how the boy was seen and cuddled.</i>
5	A	Zeig mal, wie das Huhn gefressen und geschluckt hat.	<i>Show me how the chicken ate and swallowed.</i>
5	B	Zeig mal, wie das Huhn gefressen und geschluckt wurde.	<i>Show me how the chicken was eaten and swallowed.</i>
6	A	Zeig mal, wie die Katze gejagt und gebissen hat.	<i>Show me how the cat chased and bit.</i>
6	B	Zeig mal, wie die Katze gejagt und gebissen wurde.	<i>Show me how the cat was chased and bitten.</i>
7	A	Zeig mal, wie der Junge getreten und gekniffen hat.	<i>Show me how the boy kicked and pinched.</i>
7	B	Zeig mal, wie der Junge getreten und gekniffen wurde.	<i>Show me how the boy was kicked and pinched.</i>
8	A	Zeig mal, wie der Feuerwehrmann gesucht und entdeckt hat.	<i>Show me how the fireman searched and discovered.</i>
8	B	Zeig mal, wie der Feuerwehrmann gesucht und entdeckt wurde.	<i>Show me how the fireman was searched for and discovered.</i>
9	A	Zeig mal, wie der Dieb gesucht und gesehen hat.	<i>Show me how the thief searched and saw.</i>

9	B	Zeig mal, wie der Dieb gesucht und gesehen wurde.	<i>Show me how the thief was searched for and seen.</i>
10	A	Zeig mal, wie die Maus gegessen und aufgefuttern hat.	<i>Show me how the mouse ate and ate. (synonyms)</i>
10	B	Zeig mal, wie die Maus gegessen und aufgefuttern wurde.	<i>Show me how the mouse was eaten and eaten. (synonyms)</i>
11	A	Zeig mal, wie der Stein gehauen und geschlagen hat.	<i>Show me how the rock hit and hit. (synonyms)</i>
11	B	Zeig mal, wie der Stein gehauen und geschlagen wurde.	<i>Show me how the rock was hit and hit. (synonyms)</i>
12	A	Zeig mal, wie der Junge geleckt und gebissen hat.	<i>Show me how the boy licked and bit.</i>
12	B	Zeig mal, wie der Junge geleckt und gebissen wurde.	<i>Show me how the boy was licked and bitten.</i>

Appendix 3: Critical trial materials for Experiment 4.3

Conditions:

A pre-subject active

B pre-subject passive

Target objects are as in Experiment 4.1.

Item set	Condition	German	English
1	A	Die Robbe hat gebissen und gefressen.	<i>The seal bit and ate.</i>
1	B	Die Robbe wurde gebissen und gefressen.	<i>The seal was bitten and eaten.</i>
2	A	Der Hund hat gesucht und gejagt.	<i>The dog searched and chased.</i>
2	B	Der Hund wurde gesucht und gejagt.	<i>The dog was searched for and chased.</i>
3	A	Der Frosch hat geschnappt und gefressen.	<i>The frog snapped and ate.</i>
3	B	Der Frosch wurde geschnappt und gefressen.	<i>The frog was snapped up and eaten.</i>
4	A	Der Junge hat gesehen und gestreichelt.	<i>The boy saw and cuddled.</i>
4	B	Der Junge wurde gesehen und gestreichelt.	<i>The boy was seen and cuddled.</i>
5	A	Das Huhn hat gefressen und geschluckt.	<i>The chicken ate and swallowed.</i>
5	B	Das Huhn wurde gefressen und geschluckt.	<i>The chicken was eaten and swallowed.</i>
6	A	Die Katze hat gejagt und gebissen.	<i>The cat chased and bit.</i>
6	B	Die Katze wurde gejagt und gebissen.	<i>The cat was chased and bitten.</i>
7	A	Der Junge hat getreten und gekniffen.	<i>The boy kicked and pinched.</i>
7	B	Der Junge wurde getreten und gekniffen.	<i>The boy was kicked and pinched.</i>
8	A	Der Feuerwehrmann hat gesucht und entdeckt.	<i>The fireman searched and discovered.</i>
8	B	Der Feuerwehrmann wurde gesucht und entdeckt.	<i>The fireman was searched for and discovered.</i>
9	A	Der Dieb hat gesucht und gesehen.	<i>The thief searched and saw.</i>
9	B	Der Dieb wurde gesucht und gesehen wurde.	<i>The thief was searched for and seen.</i>

10	A	Die Maus hat gegessen und aufgefuttert.	<i>The mouse ate and ate. (synonyms)</i>
10	B	Die Maus wurde gegessen und aufgefuttert.	<i>The mouse was eaten and eaten. (synonyms)</i>
11	A	Der Stein hat gehauen und geschlagen.	<i>The rock hit and hit. (synonyms)</i>
11	B	Der Stein wurde gehauen und geschlagen.	<i>The rock was hit and hit. (synonyms)</i>
12	A	Der Junge hat geleck und gebissen.	<i>The boy licked and bit.</i>
12	B	Der Junge wurde geleck und gebissen.	<i>The boy was licked and bitten.</i>

Appendix 4: Critical stimuli for Experiment 5.1

Conditions:

- A order 1; high cloze (canonical)
- B order 1; low cloze (reversed)
- C order 2; low cloze (reversed)
- D order 2; high cloze (canonical)

Context	Item	Cond.	Sentence	Target	Spillover
NNV	1	a	The aquarium visitor wondered which fish the penguins had	eaten	from the large bucket.
NNV	1	b	The aquarium visitor wondered which penguins the fish had	eaten	from the large bucket.
NNV	1	c	The aquarium visitor wondered which fish the penguins had	swum from	through the cold water.
NNV	1	d	The aquarium visitor wondered which penguins the fish had	swum from	through the cold water.
NNV	2	a	The superintendent overheard which tenant the landlord had	evicted	from the apartment.
NNV	2	b	The superintendent overheard which landlord the tenant had	evicted	from the apartment.
NNV	2	c	The superintendent overheard which tenant the landlord had	complained about	for overcharging rent.
NNV	2	d	The superintendent overheard which landlord the tenant had	complained about	for overcharging rent.
NNV	3	a	The Roman emperor asked which god the gladiator had	prayed to	before the dangerous fight.
NNV	3	b	The Roman emperor asked which gladiator the god had	prayed to	before the dangerous fight.
NNV	3	c	The Roman emperor asked which god the gladiator had	blessed	with superhuman strength.
NNV	3	d	The Roman emperor asked which gladiator the god had	blessed	with superhuman strength.
NNV	4	a	The scientist noted which antelope the lion had	killed	after a short chase.
NNV	4	b	The scientist noted which lion the antelope had	killed	after a short chase.
NNV	4	c	The scientist noted which antelope the lion had	run from	without any success.
NNV	4	d	The scientist noted which lion the antelope had	run from	without any success.
NNV	5	a	The secretary confirmed which illustrator the author had	chosen	for the novel.
NNV	5	b	The secretary confirmed which author the illustrator had	chosen	for the novel.
NNV	5	c	The secretary confirmed which illustrator the author had	drawn for	in the new children's novel.

NNV	5	d	The secretary confirmed which author the illustrator had	drawn for	in the new children's novel.
NNV	7	a	The homeowner asked which wasps the exterminator had	killed	in the local park.
NNV	7	b	The homeowner asked which exterminator the wasps had	killed	in the local park.
NNV	7	c	The homeowner asked which wasps the exterminator had	stung	in the local park.
NNV	7	d	The homeowner asked which exterminator the wasps had	stung	in the local park.
NNV	8	a	The Indians asked which buffalo the pioneers had	killed	during the stampede.
NNV	8	b	The Indians asked which pioneers the buffalo had	killed	during the stampede.
NNV	8	c	The Indians asked which buffalo the pioneers had	trampled	during the stampede.
NNV	8	d	The Indians asked which pioneers the buffalo had	trampled	during the stampede.
NNV	9	a	The historian documented which prince the assassin had	killed	at the masquerade ball.
NNV	9	b	The historian documented which assassin the prince had	killed	at the masquerade ball.
NNV	9	c	The historian documented which prince the assassin had	avoided	on the battlefield.
NNV	9	d	The historian documented which assassin the prince had	avoided	on the battlefield.
NNV	10	a	The naturalist observed which predators the deer had	avoided	in the woods.
NNV	10	b	The naturalist observed which deer the predators had	avoided	in the woods.
NNV	10	c	The naturalist observed which predators the deer had	hunted	through the woods.
NNV	10	d	The naturalist observed which deer the predators had	hunted	through the woods.
NNV	11	a	The nanny knew which housekeeper the billionaire had	hired	at a high salary.
NNV	11	b	The nanny knew which billionaire the housekeeper had	hired	at a high salary.
NNV	11	c	The nanny knew which housekeeper the billionaire had	worked for	for many years.
NNV	11	d	The nanny knew which billionaire the housekeeper had	worked for	for many years.
NNV	12	a	The park ranger documented which elephant the poacher had	killed	in the national park.
NNV	12	b	The park ranger documented which poacher the tiger had	killed	in the national park.
NNV	12	c	The park ranger documented which tiger the poacher had	mauled	in the national park.
NNV	12	d	The park ranger documented which poacher the elephant had	mauled	in the national park.
NNV	14	a	The restaurant owner forgot which customer the waitress had	served	at brunch that morning.
NNV	14	b	The restaurant owner forgot which waitress the customer had	served	at brunch that morning.
NNV	14	c	The restaurant owner forgot which customer the waitress had	tipped	at brunch that morning.
NNV	14	d	The restaurant owner forgot which waitress the customer had	tipped	at brunch that morning.
NNV	16	a	The art historian researched which model the artist had	painting	with delicate watercolors.

NNV	16	b	The art historian researched which artist the model had	Painted	with delicate watercolors.
NNV	16	c	The art historian researched which model the artist had	posed for	in the studio.
NNV	16	d	The art historian researched which artist the model had	posed for	in the studio.
NNV	17	a	The mafia boss knew which policeman the mobster had	bribed	with a suitcase of money.
NNV	17	b	The mafia boss knew which mobster the policeman had	bribed	with a suitcase of money.
NNV	17	c	The mafia boss knew which policeman the mobster had	arrested	on false charges.
NNV	17	d	The mafia boss knew which mobster the policeman had	arrested	on false charges.
NNV	18	a	The scholar wondered which king the jester had	entertained	at the royal feast.
NNV	18	b	The scholar wondered which jester the king had	entertained	at the royal feast.
NNV	18	c	The scholar wondered which king the jester had	hired	for the royal feast.
NNV	18	d	The scholar wondered which jester the king had	hired	for the royal feast.
NNV	19	a	The sports doctor recalled which player the physical therapist had	treated	after the soccer game.
NNV	19	b	The sports doctor recalled which physical therapist the player had	treated	after the soccer game.
NNV	19	c	The sports doctor recalled which player the physical therapist had	seen	after the soccer game.
NNV	19	d	The sports doctor recalled which physical therapist the player had	seen	after the soccer game.
NNV	20	a	The barkeeper explained which regulars the barmaid had	served	on Thursday night.
NNV	20	b	The barkeeper explained which barmaid the regulars had	served	on Thursday night.
NNV	20	c	The barkeeper explained which regulars the barmaid had	tipped	for excellent service.
NNV	20	d	The barkeeper explained which barmaid the regulars had	tipped	for excellent service.
NNV	21	a	The assistant guessed which skeptic the magician had	convinced	during a riveting performance.
NNV	21	b	The assistant guessed which magician the skeptic had	convinced	during a riveting performance.
NNV	21	c	The assistant guessed which skeptic the magician had	exposed	as a fraud.
NNV	21	d	The assistant guessed which magician the skeptic had	exposed	as a fraud.
NNV	22	a	The farmer's wife heard which chicken the farmhand had	killed	outside the coop.
NNV	22	b	The farmer's wife heard which farmhand the chicken had	killed	outside the coop.
NNV	22	c	The farmer's wife heard which chicken the farmhand had	pecked	outside the coop.
NNV	22	d	The farmer's wife heard which farmhand the chicken had	pecked	outside the coop.
NNV	23	a	The UFOlogist knew which astronaut the alien had	abducted	from the spaceship.
NNV	23	b	The UFOlogist knew which alien the astronaut had	abducted	from the spaceship.
NNV	23	c	The UFOlogist knew which astronaut the alien had	seen	on the ice planet.

NNV	23	d	The UFOlogist knew which alien the astronaut had	seen	on the ice planet.
NNV	24	a	The priest recorded which cannibal the missionary had	converted	in the remote territories.
NNV	24	b	The priest recorded which missionary the cannibal had	converted	in the remote territories.
NNV	24	c	The priest recorded which cannibal the missionary had	eaten	for his dinner.
NNV	24	d	The priest recorded which missionary the cannibal had	eaten	for his dinner.
NNV	25	a	The forest ranger learned which hiker the bear had	mauled	on the mountain trail.
NNV	25	b	The forest ranger learned which bear the hiker had	mauled	on the mountain trail.
NNV	25	c	The forest ranger learned which hiker the bear had	seen	in the state park.
NNV	25	d	The forest ranger learned which bear the hiker had	seen	in the state park.
NNV	26	a	The thief saw which con artist the detective had	arrested	after a long investigation.
NNV	26	b	The thief saw which detective the con artist had	arrested	after a long investigation.
NNV	26	c	The thief saw which con artist the detective had	tricked	during the interrogation.
NNV	26	d	The thief saw which detective the con artist had	tricked	during the interrogation.
NNV	27	a	The news anchor explained which criminal the journalist had	interviewed	over the phone.
NNV	27	b	The news anchor explained which journalist the criminal had	interviewed	over the phone.
NNV	27	c	The news anchor explained which criminal the journalist had	confessed to	over the phone.
NNV	27	d	The news anchor explained which journalist the criminal had	confessed to	over the phone.
NNV	28	a	The detective recalled which thief the cop had	arrested	in the bank.
NNV	28	b	The detective recalled which cop the thief had	arrested	in the bank.
NNV	28	c	The detective recalled which thief the cop had	run from	during the arrest.
NNV	28	d	The detective recalled which cop the thief had	run from	during the arrest.
NNV	29	a	The policeman observed which old lady the thief had	robbed	at a street corner.
NNV	29	b	The policeman observed which thief the old lady had	robbed	at a street corner.
NNV	29	c	The policeman observed which old lady the thief had	identified	from the line-up.
NNV	29	d	The policeman observed which thief the old lady had	identified	from the line-up.
NNV	30	a	The cattle rancher remembered which bull the cowboy had	ridden	at the local rodeo.
NNV	30	b	The cattle rancher remembered which cowboy the bull had	ridden	at the local rodeo.
NNV	30	c	The cattle rancher remembered which bull the cowboy had	gored	out on the range.
NNV	30	d	The cattle rancher remembered which cowboy the bull had	gored	out on the range.
NNV	31	a	The immunization specialist confirmed which patient the nurse had	vaccinated	for virulent measles.
NNV	31	b	The immunization specialist confirmed which nurse the patient had	vaccinated	for virulent measles.
NNV	31	c	The immunization specialist confirmed which patient the nurse had	seen	for a vaccine.

NNV	31	d	The immunization specialist confirmed which nurse the patient had	seen	for a vaccine.
NNV	32	a	The naturalist described which chick the snake had	eaten	right out of the nest.
NNV	32	b	The naturalist described which snake the chick had	eaten	right out of the nest.
NNV	32	c	The naturalist described which chick the snake had	run from	through the forest.
NNV	32	d	The naturalist described which snake the chick had	run from	through the forest.
NNV	33	a	The doctor knew which insects the biologist had	studied	in the high tech lab.
NNV	33	b	The doctor knew which biologist the insects had	studied	in the high tech lab.
NNV	33	c	The doctor knew which insects the biologist had	bitten	in the high tech lab.
NNV	33	d	The doctor knew which biologist the insects had	bitten	in the high tech lab.
NNV	34	a	The veteran forgot which war hero the president had	honored	at the grand ceremony.
NNV	34	b	The veteran forgot which president the war hero had	honored	at the grand ceremony.
NNV	34	c	The veteran forgot which war hero the president had	served	many years ago.
NNV	34	d	The veteran forgot which president the war hero had	served	many years ago.
NNV	35	a	The police detective described which bank teller the robber had	threatened	with a revolver.
NNV	35	b	The police detective described which robber the bank teller had	threatened	with a revolver.
NNV	35	c	The police detective described which bank teller the robber had	identified	after the hold-up.
NNV	35	d	The police detective described which robber the bank teller had	identified	after the hold-up.
NNV	36	a	The circus owner recalled which lion the animal trainer had	trained	for many years.
NNV	36	b	The circus owner recalled which animal trainer the lion had	trained	for many years.
NNV	36	c	The circus owner recalled which lion the animal trainer had	mauled	in the middle of the performance.
NNV	36	d	The circus owner recalled which animal trainer the lion had	mauled	in the middle of the performance.
NNV	37	a	The lab manager remembered which virus the researcher had	studied	under the microscope.
NNV	37	b	The lab manager remembered which researcher the virus had	studied	under the microscope.
NNV	37	c	The lab manager remembered which virus the researcher had	infected	after the vial broke.
NNV	37	d	The lab manager remembered which researcher the virus had	infected	after the vial broke.
NNV	38	a	The store manager noted which elderly man the IT technician had	helped	with software issues.
NNV	38	b	The store manager noted which IT technician the elderly man had	helped	with software issues.
NNV	38	c	The store manager noted which elderly man the IT technician had	asked for	earlier that day.
NNV	38	d	The store manager noted which IT technician the elderly man had	asked for	earlier that day.
NNV	39	a	The villain knew which victim the superhero had	saved	from the burning building.

NNV	39	b	The villain knew which superhero the victim had	saved	from the burning building.
NNV	39	c	The villain knew which victim the superhero had	called	during the emergency.
NNV	39	d	The villain knew which superhero the victim had	called	during the emergency.
NNV	40	a	The foreign correspondent recalled which politician the interpreter had	translated for	in the press conference.
NNV	40	b	The foreign correspondent recalled which interpreter the politician had	translated for	in the press conference.
NNV	40	c	The foreign correspondent recalled which politician the interpreter had	used	in the press conference.
NNV	40	d	The foreign correspondent recalled which interpreter the politician had	used	in the press conference.
NNV	41	a	The cybersecurity expert explained which hacker the government had	arrested	in the middle of the night.
NNV	41	b	The cybersecurity expert explained which government the hacker had	arrested	in the middle of the night.
NNV	41	c	The cybersecurity expert explained which hacker the government had	infiltrated	in the middle of the night.
NNV	41	d	The cybersecurity expert explained which government the hacker had	infiltrated	in the middle of the night.
NNV	42	a	The believer wondered which prophet the gods had	chosen	in the ancient days.
NNV	42	b	The believer wondered which gods the prophet had	chosen	in the ancient days.
NNV	42	c	The believer wondered which prophet the gods had	worshipped	in the ancient days.
NNV	42	d	The believer wondered which gods the prophet had	worshipped	in the ancient days.
NNV	43	a	The housewife knew which dog the cat had	scratched	across the nose.
NNV	43	b	The housewife knew which cat the dog had	scratched	across the nose.
NNV	43	c	The housewife knew which dog the cat had	chased	around the tree.
NNV	43	d	The housewife knew which cat the dog had	chased	around the tree.
NNV	44	a	The missionary mentioned which sinners the priest had	forgiven	during the service.
NNV	44	b	The missionary mentioned which priest the sinners had	forgiven	during the service.
NNV	44	c	The missionary mentioned which sinners the priest had	confessed to	during the service.
NNV	44	d	The missionary mentioned which priest the sinners had	confessed to	during the service.
NNV	45	a	The mother watched which swans the child had	fed	at the pond.
NNV	45	b	The mother watched which child the swans had	fed	at the pond.
NNV	45	c	The mother watched which swans the child had	swum to	at the pond.
NNV	45	d	The mother watched which child the swans had	swum to	at the pond.
NNV	46	a	The biologist recorded which mouse the owl had	eaten	after the drought.
NNV	46	b	The biologist recorded which owl the mouse had	eaten	after the drought.
NNV	46	c	The biologist recorded which mouse the owl had	avoided	after the drought.
NNV	46	d	The biologist recorded which owl the mouse had	avoided	after the drought.
NNV	47	a	The riding instructor remembered which pony the girl had	ridden	for the first time.
NNV	47	b	The riding instructor remembered which girl the pony had	ridden	for the first time.
NNV	47	c	The riding instructor remembered which pony the girl had	thrown off	for the first time.
NNV	47	d	The riding instructor remembered which girl the	thrown off	for the first time.

			pony had		
NNV	48	a	The board of trustees noted which fraudster the tax agent had	caught	during the audit.
NNV	48	b	The board of trustees noted which tax agent the fraudster had	caught	during the audit.
NNV	48	c	The board of trustees noted which fraudster the tax agent had	swindled	during the audit.
NNV	48	d	The board of trustees noted which tax agent the fraudster had	swindled	during the audit.
NNV	49	a	The game show host recalled which cheater the judge had	disqualified	in the final round.
NNV	49	b	The game show host recalled which judge the cheater had	disqualified	in the final round.
NNV	49	c	The game show host recalled which cheater the judge had	bribed	before the show.
NNV	49	d	The game show host recalled which judge the cheater had	bribed	before the show.
NNV	50	a	The historian researched which traitor the king had	executed	during the uprising.
NNV	50	b	The historian researched which king the traitor had	executed	during the uprising.
NNV	50	c	The historian researched which traitor the king had	betrayed	during the uprising.
NNV	50	d	The historian researched which king the traitor had	betrayed	during the uprising.
NNV	51	a	The psychic understood which teenagers the ghost had	haunted	in the abandoned mansion.
NNV	51	b	The psychic understood which ghost the teenagers had	haunted	in the abandoned mansion.
NNV	51	c	The psychic understood which teenagers the ghost had	seen	in the abandoned mansion.
NNV	51	d	The psychic understood which ghost the teenagers had	seen	in the abandoned mansion.
NNV	52	a	The trapper heard which buck the hunter had	shot	in the forest.
NNV	52	b	The trapper heard which hunter the buck had	shot	in the forest.
NNV	52	c	The trapper heard which buck the hunter had	run from	in the forest.
NNV	52	d	The trapper heard which hunter the buck had	run from	in the forest.
NNV	53	a	The naturalist observed which snake the scorpion had	stung	in the desert cave.
NNV	53	b	The naturalist observed which scorpion the snake had	stung	in the desert cave.
NNV	53	c	The naturalist observed which snake the scorpion had	eaten	in the desert cave.
NNV	53	d	The naturalist observed which scorpion the snake had	eaten	in the desert cave.
NNV	54	a	The foreign correspondent reported which dictator the activist had	protested	outside the capitol building.
NNV	54	b	The foreign correspondent reported which activist the dictator had	protested	outside the capitol building.
NNV	54	c	The foreign correspondent reported which dictator the activist had	imprisoned	for thirty years.
NNV	54	d	The foreign correspondent reported which activist the dictator had	imprisoned	for thirty years.
NNV	55	a	The bartender saw which teenager the bouncer had	thrown out	from the crowded club.
NNV	55	b	The bartender saw which bouncer the teenager had	thrown out	from the crowded club.

			had		
NNV	55	c	The bartender saw which teenager the bouncer had	bribed	at the entrance.
NNV	55	d	The bartender saw which bouncer the teenager had	bribed	at the entrance.
NNV	56	a	The highway patrolman confirmed which elderly woman the hitchhiker had	robbed	on the side of the road.
NNV	56	b	The highway patrolman confirmed which hitchhiker the elderly woman had	robbed	on the side of the road.
NNV	56	c	The highway patrolman confirmed which elderly woman the hitchhiker had	hit	on the side of the road.
NNV	56	d	The highway patrolman confirmed which hitchhiker the elderly woman had	hit	on the side of the road.
NNV	57	a	The shepherd forgot which sheep the wolf had	eaten	from the flock.
NNV	57	b	The shepherd forgot which wolf the sheep had	eaten	from the flock.
NNV	57	c	The shepherd forgot which sheep the wolf had	butted	away from the flock.
NNV	57	d	The shepherd forgot which wolf the sheep had	butted	away from the flock.
NNV	109	a	The stable boy remembered which mule the trader had	sold	at the busy market.
NNV	109	b	The stable boy remembered which trader the mule had	sold	at the busy market.
NNV	109	c	The stable boy remembered which mule the trader had	kicked	at the busy market.
NNV	109	d	The stable boy remembered which trader the mule had	kicked	at the busy market.
NNV	111	a	The insect collector observed which spider the bee had	stung	in the garden.
NNV	111	b	The insect collector observed which bee the spider had	stung	in the garden.
NNV	111	c	The insect collector observed which spider the bee had	trapped	in the web.
NNV	111	d	The insect collector observed which bee the spider had	trapped	in the web.
NNV	114	a	The aquarium owner knew which orca the animal rights activist had	freed	in the middle of the night.
NNV	114	b	The aquarium owner knew which animal rights activist the orca had	freed	in the middle of the night.
NNV	114	c	The aquarium owner knew which orca the animal rights activist had	befriended	after many visits.
NNV	114	d	The aquarium owner knew which animal rights activist the orca had	befriended	after many visits.
NNV	115	a	The principal recorded which stoner the snitch had	reported	during an assembly.
NNV	115	b	The principal recorded which snitch the stoner had	reported	during an assembly.
NNV	115	c	The principal recorded which stoner the snitch had	beaten up	in the locker room.
NNV	115	d	The principal recorded which snitch the stoner had	beaten up	in the locker room.
NNV	116	a	The butler remembered which chauffeur the millionaire had	hired	the previous week.
NNV	116	b	The butler remembered which millionaire the chauffeur had	hired	the previous week.
NNV	116	c	The butler remembered which chauffeur the millionaire had	driven	the previous week.
NNV	116	d	The butler remembered which millionaire the	driven	the previous week.

			chauffeur had		
NNV	117	a	The farmhand observed which cow the farmer had	milked	in the big red barn.
NNV	117	b	The farmhand observed which farmer the cow had	milked	in the big red barn.
NNV	117	c	The farmhand observed which cow the farmer had	kicked	in the big red barn.
NNV	117	d	The farmhand observed which farmer the cow had	kicked	in the big red barn.
NVN	1	a	The queen wondered which maid had dressed the	princess	before the great feast.
NVN	1	b	The queen wondered which princess had dressed the	maid	before the great feast.
NVN	2	a	The ringmaster noticed which clown had teased the	children	with hilarious tricks and jokes.
NVN	2	b	The ringmaster noticed which children had teased the	clown	with hilarious tricks and jokes.
NVN	3	a	The plantation owner heard which soldier had hidden the	slave	on the journey north.
NVN	3	b	The plantation owner heard which slave had hidden the	soldier	on the journey north.
NVN	4	a	The corporal knew which troops had deposed the	dictator	in the military coup.
NVN	4	b	The corporal knew which dictator had deposed the	troops	in the military coup.
NVN	5	a	The midwife indicated which mother had trusted the	doctor	at the private hospital.
NVN	5	b	The midwife indicated which doctor had trusted the	mother	at the private hospital.
NVN	6	a	The Nobel committee asked which scientist had mentored the	student	at the university.
NVN	6	b	The Nobel committee asked which student had mentored the	scientist	at the university.
NVN	7	a	The horse trainer saw which jockey had raced the	horse	across the track.
NVN	7	b	The horse trainer saw which horse had raced the	jockey	across the track.
NVN	8	a	The colonist recalled which witch had charmed the	villagers	with a wicked spell.
NVN	8	b	The colonist recalled which villagers had charmed the	witch	with a wicked spell.
NVN	9	a	The social worker explained which foster parent had raised the	child	from a young age.
NVN	9	b	The social worker explained which child had raised the	foster parent	from a young age.
NVN	10	a	The monk documented which native had welcomed the	priest	to the remote village.
NVN	10	b	The monk documented which priest had welcomed the	native	to the remote village.
NVN	11	a	The janitor guessed which teacher had teased the	kid	about passing notes in class.
NVN	11	b	The janitor guessed which kid had teased the	teacher	about passing notes in class.
NVN	12	a	The cotton farmer forgot which master had whipped the	slave	in the field.
NVN	12	b	The cotton farmer forgot which slave had whipped the	master	in the field.
NVN	13	a	The groom learned which groomsman had hired the	stripper	for the party.
NVN	13	b	The groom learned which stripper had hired the	groomsman	for the party.

NVN	14	a	The mother heard which babysitter had fed the	toddler	early in the morning.
NVN	14	b	The mother heard which toddler had fed the	babysitter	early in the morning.
NVN	15	a	The receptionist noted which doctor had examined the	patient	in the private room.
NVN	15	b	The receptionist noted which patient had examined the	doctor	in the private room.
NVN	16	a	The soup kitchen owner remembered which volunteer had sheltered the	homeless person	during the storm.
NVN	16	b	The soup kitchen owner remembered which homeless person had sheltered the	volunteer	during the storm.
NVN	17	a	The secretary overheard which boss had promoted the	employee	for hard work.
NVN	17	b	The secretary overheard which employee had promoted the	boss	for hard work.
NVN	18	a	The tabloid reporter publicized which photographer had followed the	celebrity	after the awards show.
NVN	18	b	The tabloid reporter publicized which celebrity had followed the	photographer	after the awards show.
NVN	19	a	The death row guard recorded which judge had blamed the	criminal	during the trial.
NVN	19	b	The death row guard recorded which criminal had blamed the	judge	during the trial.
NVN	20	a	The cowboy learned which sheriff had caught the	thief	in the local saloon.
NVN	20	b	The cowboy learned which thief had caught the	sheriff	in the local saloon.
NVN	21	a	The game show host saw which audience member had congratulated the	winner	after the finale.
NVN	21	b	The game show host saw which winner had congratulated the	audience member	after the finale.
NVN	22	a	The churchgoer remembered which demon had possessed the	priest	during the sermon.
NVN	22	b	The churchgoer remembered which priest had possessed the	demon	during the sermon.
NVN	23	a	The cheerleader saw which nerd had envied the	quarterback	during gym class.
NVN	23	b	The cheerleader saw which quarterback had envied the	nerd	during gym class.
NVN	24	a	The bishop recorded which saint had healed the	leper	in the slums.
NVN	24	b	The bishop recorded which leper had healed the	saint	in the slums.
NVN	25	a	The policeman recalled which victim had escaped the	murderer	through the open window.
NVN	25	b	The policeman recalled which murderer had escaped the	victim	through the open window.
NVN	26	a	The prosecutor noted which lawyer had defended the	criminal	for the high profile case.
NVN	26	b	The prosecutor noted which criminal had defended the	lawyer	for the high profile case.
NVN	27	a	The campaign manager observed which audience had supported the	candidate	during the debate.
NVN	27	b	The campaign manager observed which candidate had supported the	audience	during the debate.
NVN	28	a	The tour manager noticed which crowd had idolized the	singer	at the concert.
NVN	28	b	The tour manager noticed which singer had idolized the	crowd	at the concert.
NVN	29	a	The housekeeper observed which cat had devoured the	mouse	in the garage.
NVN	29	b	The housekeeper observed which mouse had	cat	in the garage.

			devoured the		
NVN	30	a	The gossip columnist learned which paparazzi had followed the	movie star	during a beach vacation.
NVN	30	b	The gossip columnist learned which movie star had followed the	paparazzi	during a beach vacation.
NVN	31	a	The rodeo clown knew which bull had trampled the	cowboy	the previous afternoon.
NVN	31	b	The rodeo clown knew which cowboy had trampled the	bull	the previous afternoon.
NVN	32	a	The beachgoer saw which lifeguard had rescued the	swimmer	at high tide.
NVN	32	b	The beachgoer saw which swimmer had rescued the	lifeguard	at high tide.
NVN	33	a	The jeweler remembered which policeman had grabbed the	thief	with a firm grip.
NVN	33	b	The jeweler remembered which thief had grabbed the	policeman	with a firm grip.
NVN	34	a	The bouncer confirmed which heckler had confused the	comedian	at the open mic night.
NVN	34	b	The bouncer confirmed which comedian had confused the	heckler	at the open mic night.
NVN	35	a	The military historian reported which enemy had overthrown the	government	with a surprise attack.
NVN	35	b	The military historian reported which government had overthrown the	rebels	with a surprise attack.
NVN	36	a	The storyteller explained which princess had feared the	dragon	for a long time.
NVN	36	b	The storyteller explained which dragon had feared the	princess	for a long time.
NVN	37	a	The baron speculated which prince had betrayed the	king	at the tournament.
NVN	37	b	The baron speculated which king had betrayed the	prince	at the tournament.
NVN	38	a	The bridal shop owner mentioned which employee had comforted the	bride	in the dressing room.
NVN	38	b	The bridal shop owner mentioned which bride had comforted the	employee	in the dressing room.
NVN	39	a	The high schooler explained which teacher had bullied the	kid	during evening detention.
NVN	39	b	The high schooler explained which kid had bullied the	teacher	during evening detention.
NVN	40	a	The CIA agent knew which pilot had fought the	terrorists	with air attacks.
NVN	40	b	The CIA agent knew which terrorists had fought the	pilot	with air attacks.
NVN	41	a	The safari guide speculated which hunter had stalked the	lion	across the savannah.
NVN	41	b	The safari guide speculated which lion had stalked the	hunter	across the savannah.
NVN	42	a	The United Nations delegate reported which terrorist had murdered the	president	at the embassy.
NVN	42	b	The United Nations delegate reported which president had murdered the	terrorist	at the embassy.
NVN	43	a	The movie producer noted which director had fired the	actor	without any warning.
NVN	43	b	The movie producer noted which actor had fired the	director	without any warning.
NVN	44	a	The zoo director recalled which zookeeper had	lion	inside the enclosure.

		fed the		
NVN	44	b	The zoo director recalled which lion had fed the	zookeeper
			The first mate remembered which whale had	
NVN	45	a	swallowed the	sailor
			The first mate remembered which sailor had	
NVN	45	b	swallowed the	whale
			The junkie observed which policeman had chased	
NVN	46	a	the	drug dealer
			The junkie observed which drug dealer had	
NVN	46	b	chased the	policeman
			The book reviewer speculated which publisher	
NVN	47	a	had interviewed the	author
			The book reviewer speculated which author had	
NVN	47	b	interviewed the	publisher
			The mental health worker documented which	
NVN	48	a	psychologist had analyzed the	patient
			The mental health worker documented which	
NVN	48	b	patient had analyzed the	psychologist
			The mayor remembered which policeman had	
NVN	49	a	rescued the	child
			The mayor remembered which child had rescued	
NVN	49	b	the	policeman
			The town scribe recorded which ghost had scared	
NVN	50	a	the	child
			The town scribe recorded which child had scared	
NVN	50	b	the	ghost
			The werewolf figured out which vampire had	
NVN	51	a	bitten the	girl
			The werewolf figured out which girl had bitten	
NVN	51	b	the	vampire
			The party planner indicated which host had	
NVN	52	a	invited the	guest
			The party planner indicated which guest had	
NVN	52	b	invited the	host
			The juror noticed which attorney had questioned	
NVN	53	a	the	witness
			The juror noticed which witness had questioned	
NVN	53	b	the	attorney
			The congressman publicized which terrorist had	
NVN	54	a	assassinated the	politician
			The congressman publicized which politician had	
NVN	54	b	assassinated the	terrorist
			The milkman confirmed which mailman had	
NVN	55	a	escaped the	dog
			The milkman confirmed which dog had escaped	
NVN	55	b	the	mailman
			The lifeguard speculated which shark had eaten	
NVN	56	a	the	surfer
			The lifeguard speculated which surfer had eaten	
NVN	56	b	the	shark
			The storyteller explained which mermaid had	
NVN	57	a	seduced the	prince
			The storyteller explained which prince had	
NVN	57	b	seduced the	mermaid
			The aquarium worker explained which dolphin	
NVN	58	a	had obeyed the	trainer

NVN	58	b	The aquarium worker explained which trainer had obeyed the	dolphin	during the routine.
NVN	59	a	The cult leader knew which demon had frightened the	follower	during the ceremony.
NVN	59	b	The cult leader knew which follower had frightened the	demon	during the ceremony.
NVN	60	a	The football commentator noted which coach had trained the	player	since high school.
NVN	60	b	The football commentator noted which player had trained the	coach	since high school.

References

- Abbot-Smith, K., & Behrens, H. (2006). How known constructions influence the acquisition of other constructions: the German passive and future constructions. *Cognitive Science*, 30(6), 995–1026.
http://doi.org/10.1207/s15516709cog0000_61
- Alcock, K. J., Rimba, K., & Newton, C. R. J. C. (2012). Early production of the passive in two Eastern Bantu languages. *First Language*, 32(4), 459–478.
<http://doi.org/10.1177/0142723711419328>
- Allen, S. E. M., & Crago, M. B. (1996). Early passive acquisition in Inuktitut. *Journal of Child Language*, 23(September), 129–155.
<http://doi.org/10.1017/S0305000900010126>
- Amt für Statistik Berlin-Brandenburg. (2015). *Ergebnisse des Mikrozensus im Land Berlin 2015 – Haushalte, Familien und Lebensformen*. Retrieved from https://www.statistik-berlin-brandenburg.de/publikationen/stat_berichte/2016/SB_A01-11-00_2015j01_BE.pdf
- Armon-Lotem, S., Haman, E., Jensen de López, K., Smoczyńska, M., Yatsushiro, K., Szczerbinski, M., ... van der Lely, H. (2016). A large-scale cross-linguistic investigation of the acquisition of passive. *Language Acquisition*, 23(1), 27–56.
<http://doi.org/10.1080/10489223.2015.1047095>
- Aschermann, E., Gülzow, I., & Wendt, D. (2004). Differences in the Comprehension of Passive Voice in German- and. *Swiss Journal of Psychology*, 63(4), 235–245.

<http://doi.org/10.1024/1421-0185.63.4.235>

- Bader, M., & Häussler, J. (2010). Word order in German: A corpus study. *Lingua*, *120*, 717–762. <http://doi.org/10.1016/j.lingua.2009.05.007>
- Bader, M., & Meng, M. (1999). Subject-Object Ambiguities in German Embedded Clauses: An Across-the-Board Comparison. *Journal of Psycholinguistic Research*, *28*(2), 121–143.
- Bartke, S. (2004). Passives in German children with Williams Syndrome. In S. Bartke & J. Siegmüller (Eds.), *Williams Syndrome across Languages* (pp. 345–370). Amsterdam: Benjamins.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.
- Bates, E., & MacWhinney, B. Competition, Variation, and Language Learning (1989).
- Behrens, H. (2006). The input-output relationship in first language acquisition. *Language and Cognitive Processes*, *21*, 2–24.
- Bever, T. (1970). The Cognitive Basis for Linguistic Structures. In R. Hayes (Ed.), *Cognition and Language Development* (pp. 279–362). New York: Wiley and Sons.
- Borer, H., & Wexler, K. (1987). The maturation of syntax. In H. Borer, K. Wexler, T. Roeper, & E. Williams (Eds.), *Parameter Setting* (pp. 123–172).
- Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., ... Schlewsky, M. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and*

- Language*, 117(3), 133–152. <http://doi.org/10.1016/j.bandl.2010.09.010>
- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*, 59(1), 55–73. <http://doi.org/10.1016/j.brainresrev.2008.05.003>
- Bornkessel, I., & Schlewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113(4), 787–821. <http://doi.org/10.1037/0033-295X.113.4.787>
- Bornkessel, I., Schlewsky, M., & Friederici, A. D. (2002). Grammar overrides frequency: evidence from the online processing of flexible word order. *Cognition*, 85, 21–30.
- Brown, C. M., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, 5(1), 34–44.
- Choi, Y., & Trueswell, J. C. (2010). Children’s (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, 106(1), 41–61. <http://doi.org/10.1016/j.jecp.2010.01.003>
- Chow, W.-Y., Lau, E., Wang, S., & Phillips, C. (2018). Wait a second! Delayed impact of argument roles on on-line verb prediction. *Language, Cognition and Neuroscience*.
- Chow, W.-Y., Momma, S., Smith, C., Lau, E., & Phillips, C. (2016). Prediction as memory retrieval: timing and mechanisms. *Language, Cognition and Neuroscience*, 31(5), 617–627. <http://doi.org/10.1080/23273798.2016.1160135>

- Chow, W.-Y., & Phillips, C. (2013). No semantic illusions in the “Semantic P600” phenomenon: ERP evidence from Mandarin Chinese. *Brain Research, 1506*, 76–93. <http://doi.org/10.1016/j.brainres.2013.02.016>
- Chow, W.-Y., Smith, C., Lau, E., & Phillips, C. (2015). A ‘bag-of-arguments’ mechanism for initial verb predictions. *Language, Cognition and Neuroscience, 44*(September). <http://doi.org/10.1080/23273798.2015.1066832>
- Christianson, K. (2002). *Sentence processing in a “nonconfigurational” language*. ProQuest Dissertations and Theses. University of Michigan. Retrieved from http://ezproxy.net.ucf.edu/login?url=http://search.proquest.com/docview/305599947?accountid=10003%5Cnhttp://sfx.fcla.edu/ucf?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+&theses&sid=ProQ:ProQuest+Dissertations+&T
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic Roles Assigned along the Garden Path Linger. *Cognitive Psychology, 42*, 368–407. <http://doi.org/10.1006/cogp.2001.0752>
- Das Deutsche Referenzkorpus DeReKo. (n.d.). Mannheim, Institut für deutsche Sprache.
- de Villiers, J. G., Flusberg, H. B. T., Hakuta, K., & Cohen, M. (1979). Children’s Comprehension of Relative Clauses. *Journal of Psycholinguistic Research, 8*(5), 499–518.
- Demiral, Ş. B., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2008). On the universality of language comprehension strategies: Evidence from Turkish. *Cognition, 106*(1), 484–500. <http://doi.org/10.1016/j.cognition.2007.01.008>

- Demuth, K. (1989). Maturation and the Acquisition of the Sesotho Passive. *Language*, 65(1), 56–80.
- Demuth, K. (1990). Subject, topic and Sesotho passive. *Journal of Child Language*, 17(1), 67–84.
- Demuth, K., Moloi, F., & Machobane, M. (2010). 3-Year-olds' comprehension, production, and generalization of Sesotho passives. *Cognition*, 115(2), 238–251. <http://doi.org/10.1016/j.cognition.2009.12.015>
- Diessel, H., & Tomasello, M. (2005). A New Look at the Acquisition of Relative Clauses. *Language*, 81(4), 882–906.
- Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2014). Familiar verbs are not always easier than novel verbs: How German pre-school children comprehend active and passive sentences. *Cognitive Science*, 38(1), 128–151. <http://doi.org/10.1111/cogs.12066>
- Dittmar, M., Abbott-Smith, K., Lieven, E., & Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child Development*, 79(4), 1152–1167.
- 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.
- Engelhardt, P. E. (2014). Children's and Adolescents' Processing of Temporary Syntactic Ambiguity: An Eye Movement Study. *Child Development Research*, 2014, 1–13.
- Ettinger, A. (2018). *Relating lexical and syntactic processes in language: bridging research in humans and machines*. University of Maryland, College Park.

- Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name : Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, *41*, 469–495. <http://doi.org/10.1006/jmla.1999.2660>
- Felser, C., Clahsen, H., & Münte, T. F. (2003). Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and wh-movement in German. *Brain and Language*, *87*(3), 345–354. [http://doi.org/10.1016/S0093-934X\(03\)00135-4](http://doi.org/10.1016/S0093-934X(03)00135-4)
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*, 164–203. [http://doi.org/10.1016/S0010-0285\(03\)00005-7](http://doi.org/10.1016/S0010-0285(03)00005-7)
- Fox, D., & Grodzinsky, Y. (1998). Children’s Passive: A View from the By-Phrase. *Linguistic Inquiry*, *29*(2), 311–332.
- Frazier, L. (1987). Syntactic processing: evidence from Dutch. *Natural Language & Linguistic Theory*, *5*(4), 519–559.
- Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.
- Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The Contributions of Verb Bias and Plausibility to the Comprehension of Temporarily Ambiguous Sentences. *Journal of Memory and Language*, *37*, 58–93.
- Gordon, P., & Chafetz, J. (1990). Verb-based versus class-based accounts of actionality in children’s comprehension of passives. *Cognition*, *36*, 227–254.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438–441.

- Harris, M. (1976). The influence of reversibility and truncation on the interpretation of the passive voice by young children. *British Journal of Psychology*, 67(3), 419–427.
- Hemforth, B., & Konieczny, L. (2000). *German Sentence Processing*. Dordrecht: Kluwer Academic.
- Hirsch, C., & Wexler, K. (2006). By the Way, Children Don't Know By.
- Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Brain Research. Cognitive Brain Research*, 19(1), 59–73.
<http://doi.org/10.1016/j.cogbrainres.2003.10.022>
- Huang, Y. T., Leech, K., & Rowe, M. L. (2017). Exploring socioeconomic differences in syntactic development through the lens of real-time processing. *Cognition*, 159, 61–75. <http://doi.org/10.1016/j.cognition.2016.11.004>
- Huang, Y. T., Zheng, X., Meng, X., & Snedeker, J. (2013). Children's assignment of grammatical roles in the online processing of Mandarin passive sentences. *Journal of Memory and Language*, 69(4), 589–606.
<http://doi.org/10.1016/j.jml.2013.08.002>
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, 99(1), 122–149. <http://doi.org/10.1037/0033-295X.99.1.122>
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1), 133–156.

[http://doi.org/10.1016/S0749-596X\(03\)00023-8](http://doi.org/10.1016/S0749-596X(03)00023-8)

Kamide, Y., & Mitchell, D. C. (1999). Incremental Pre-head Attachment in Japanese Parsing. *Language and Cognitive Processes*, *14*(5–6), 631–662.

<http://doi.org/10.1080/016909699386211>

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*(1), 37–55.

Kidd, E., Brandt, S., Lieven, E., Tomasello, M., & Kidd, E. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses children's processing of relative clauses, *0965*(December). <http://doi.org/10.1080/01690960601155284>

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225. <http://doi.org/10.1016/j.jml.2004.10.002>

Kim, A., & Sikos, L. (2011). Conflict and surrender during sentence processing: An ERP study of syntax-semantics interaction. *Brain and Language*, *118*(1–2), 15–22.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition*, *95*, 95–127. <http://doi.org/10.1016/j.cognition.2004.03.002>

Knoeferle, P., Habets, B., Crocker, M. W., & Münte, T. F. (2008). Visual Scenes Trigger Immediate Syntactic Reanalysis : Evidence from ERPs during Situated

- Spoken Comprehension. *Cerebral Cortex*, 18(April), 789–795.
<http://doi.org/10.1093/cercor/bhm121>
- Kolk, H. H. J., Chwilla, D. J., Van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, 85(1), 1–36. [http://doi.org/10.1016/S0093-934X\(02\)00548-5](http://doi.org/10.1016/S0093-934X(02)00548-5)
- Konieczny, L., Hemforth, B., Scheepers, C., Strube, G., Konieczny, L., Hemforth, B., ... Strube, G. (1997). The Role of Lexical Heads in Parsing: Evidence from German. *Language and Cognitive Processes*, 12(2/3), 307–348.
- Kukona, A., Fang, S. Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, 119(1), 23–42.
<http://doi.org/10.1016/j.cognition.2010.12.002>
- Kuperberg, G. (2015). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*.
- Kuperberg, G., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59.
<http://doi.org/10.1080/23273798.2015.1102299>
- Kuperberg, G., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17(1), 117–129.
[http://doi.org/10.1016/S0926-6410\(03\)00086-7](http://doi.org/10.1016/S0926-6410(03)00086-7)
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials

- reflect semantic incongruity. *Science (New York, N.Y.)*.
<http://doi.org/10.1126/science.7350657>
- Kutas, M., & Hillyard, S. A. (1984). An Electrophysiological Probe of Incidental Semantic Association. *Journal of Cognitive Neuroscience*, *1*(1), 38–49.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26.
- Lau, E. F., Holcomb, P. J., & Kuperberg, G. (2013). Dissociating N400 Effects of Prediction from Association in Single-word Contexts. *Journal of Cognitive Neuroscience*, *25*(3), 484–502. http://doi.org/10.1162/jocn_a_00328
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933.
<http://doi.org/10.1038/nrn2532>
- Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–198.
- Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464. <http://doi.org/10.1016/j.jml.2010.07.003>
- Lewis, S., & Phillips, C. (2015). Aligning Grammatical Theories and Language Processing Models. *Journal of Psycholinguistic Research*, *44*(1), 27–46.
<http://doi.org/10.1007/s10936-014-9329-z>
- Lindner, K. (2003). The development of sentence-interpretation strategies in

- monolingual German-learning children with and without specific language impairment. *Linguistics*, 41(2), 213–254. <http://doi.org/10.1515/ling.2003.008>
- MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue Validity and Sentence Interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, 23, 127–150.
- Mani, N., & Huettig, F. (2012). Prediction During Language Processing is a Piece of Cake — But Only for Skilled Producers. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4), 843–847.
<http://doi.org/10.1037/a0029284>
- Maratsos, M., Fox, D. E. C., Becker, J. A., & Chalkley, M. A. (1985). Semantic restrictions on children's passives. *Cognition*, 19(2), 167–191.
- Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead : Information-processing time with and without saccades. *Perception & Psychophysics*, 53(4), 372–380.
- Matzke, M., Mai, H., Nager, W., Rüsseler, J., & Münte, T. F. (2002). The costs of freedom: an ERP – study of non-canonical sentences. *Clinical Neurophysiology*, 113, 844–852.
- Mazuka, R., Jincho, N., & Oishi, H. (2009). Development of Executive Control and Language Processing. *Language and Linguistics Compass*, 3(1), 59–89.
<http://doi.org/10.1111/j.1749-818X.2008.00102.x>
- McRae, K., & Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Linguistics and Language Compass*, 3(6), 1417–1429. <http://doi.org/10.1111/j.1749-818X.2009.00174.x>

- Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions: an analysis with event-related potentials. *Memory & Cognition*, 23(4), 477–494.
<http://doi.org/10.3758/BF03197249>
- Messenger, K., Branigan, H. P., & McLean, J. F. (2012). Is children's acquisition of the passive a staged process? Evidence from six- and nine-year-olds' production of passives. *Journal of Child Language*, 39(05), 991–1016.
<http://doi.org/10.1017/S0305000911000377>
- Mills, A. E. (1985). The acquisition of German. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition* (pp. 141–254). Hillsdale, NJ: Lawrence Erlbaum.
- Momma, S. (2016). *Parsing, Generation, and Grammar*. University of Maryland, College Park.
- Nation, K., Marshall, C. M., & Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology*, 86(4), 314–329. <http://doi.org/10.1016/j.jecp.2003.09.001>
- Novick, J. M., Kan, I. P., & Thompson-Schill, S. L. (2009). *A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex* (Vol. 26).
<http://doi.org/10.1080/02643290903519367.A>
- 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.
<http://doi.org/10.3758/CABN.5.3.263>
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.
- Philipp, M., Bornkessel-Schlesewsky, I., Bisang, W., & Schlewsky, M. (2008). The role of animacy in the real time comprehension of Mandarin Chinese: Evidence from auditory event-related brain potentials. *Brain and Language*, 105, 112–133.
<http://doi.org/10.1016/j.bandl.2007.09.005>
- Pinker, S., Lebeaux, D. S., & Frost, L. A. (1987). Productivity and constraints in the acquisition of the passive. *Cognition*, 26, 195–267.
- Poeppel, D., & Wexler, K. (1993). The Full Competence Hypothesis of Clause Structure in Early German. *Language*, 69(1), 1–33.
- Potsdam, A. für S. (2015). *Wirtschaft & Statistik Potsdam*.
- Pye, C. (1991). The Acquisition of K'iche' (Maya). In D. I. Slobin (Ed.), *The Cross-Linguistic Study of Language Acquisition* (Vol. 3, pp. 221–308). Hillsdale, NJ: Erlbaum.
- Pye, C., & Poz, P. Q. (1988). Precocious Passives (and Antipassives) in Quiche Mayan. *Papers and Reports on Child Language Development*, 27(9), 1689–1699. <http://doi.org/10.1017/CBO9781107415324.004>
- R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, Vienna, Austria.
Retrieved from <https://www.r-project.org/>
- Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical

- structures: A corpus analysis. *Journal of Memory and Language*, 57, 348–379.
<http://doi.org/10.1016/j.jml.2007.03.002>
- Schaner-Wolles, C. (1989). Strategies in acquiring grammatical relations in German: Word order or case marking. *Folia Linguistica*, 23(1–2), 131–156.
- Schipke, C. S., Friederici, A. D., & Oberecker, R. (2011). Brain responses to case-marking violations in German preschool children. *Developmental Neuroscience*, 22(16), 850–854. <http://doi.org/10.1097/WNR.0b013e32834c1578>
- Schipke, C. S., Knoll, L. J., Friederici, A. D., & Oberecker, R. (2012). Preschool children's interpretation of object-initial sentences: Neural correlates of their behavioral performance. *Developmental Science*, 15(6), 762–774.
<http://doi.org/10.1111/j.1467-7687.2012.01167.x>
- Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2009). When semantic P600s turn into N400s: On cross-linguistic differences in online verb-argument linking. In M. Horne, M. Lindgren, M. Roll, K. Alter, & J. von Koss Torkildsen (Eds.), *Brain Talk: Discourse with and in the brain. Papers from the first Birgit Rausing Language Program Conference in Linguistics* (pp. 75–97). Lund: Birgit Rausing Language Program.
- Schlesewsky, M., Fanselow, G., Kliegl, R., & Krems, J. (2000). The subject preference in the processing of locally ambiguous wh-questions in German. In B. Hemforth & L. Konieczny (Eds.), *German Sentence Processing* (pp. 65–93). Dordrecht: Kluwer Academic.
- Schriefers, H., Friederici, A. D., & Kühn, K. (1995). The Processing of Locally Ambiguous Relative Clauses in German. *Journal of Memory and Language*, 34,

499–520.

- Sekerina, I., Stromswold, K., & Hestvik, A. (2004). How do adults and children process referentially ambiguous pronouns? *Journal of Child Language*, *31*(1), 123–152. <http://doi.org/10.1017/S0305000903005890>
- Sereno, S. C., Pacht, J. M., & Rayner, K. (1991). The effect of meaning frequency on processing lexically ambiguous words: Evidence from eye fixations. *Psychological Science*, *3*(5), Psychol. Sci.
- Shannon, C. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(July & October), 379–423 & 623–656.
- Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingered misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, *69*(2), 104–120. <http://doi.org/10.1016/j.jml.2013.04.001>
- Smith, N., & Levy, R. P. (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*(33).
- Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, *82*, 1–17. <http://doi.org/10.1016/j.jml.2015.02.004>
- Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1280–1293.

- Strohner, H., & Nelson, K. E. (1974). The Young Child's Development of Sentence Comprehension: Influence of Event Probability, Nonverbal Context, Syntactic Form, and Strategies. *Child Development*, 45(3), 567–576.
- Stroud, C., & Phillips, C. (2012). Examining the evidence for an independent semantic analyzer: An ERP study in Spanish. *Brain and Language*, 120(2), 108–126. <http://doi.org/10.1016/j.bandl.2011.02.001>
- Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, 105(2), 477–488. <http://doi.org/10.1016/j.cognition.2006.10.009>
- Tanenhaus, M. K., Carlson, G. N., & Trueswell, J. C. (1989). The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes*, 4(3–4), 211–234.
- Taylor, W. L. (1953). “Cloze procedure”: A new tool for measuring readability. *Journalism Bulletin*, 30(4), 415–433.
- Torchiano, M. (2017). *effsize: Efficient Effect Size Computation*.
- Trueswell, J. C., & Gleitman, L. R. (2004). Trueswell, John, and Lila Gleitman. "Children's eye movements during listening: Developmental evidence for a constraint-based theory of sentence processing. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 319–346). New York, NY: Psychology Press.
- Trueswell, J. C., Kaufman, D., Hafri, A., & Lidz, J. (2012). Development of parsing abilities interacts with grammar learning: Evidence from Tagalog and Kannada. *Proceedings of BUCLD 36*.
- Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-

- path effect: Studying on-line sentence processing in young children. *Cognition*, 73(2), 89–134. [http://doi.org/10.1016/S0010-0277\(99\)00032-3](http://doi.org/10.1016/S0010-0277(99)00032-3)
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*. <http://doi.org/10.1006/jmla.1994.1014>
- Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(3), 528–553.
- Turner, E. A., & Rommetveit, R. (1967). The Acquisition of Sentence Voice and Reversibility. *Child Delevopment*, 38(3), 649–660.
- Ueno, M., & Kluender, R. (2003). Event-related brain indices of Japanese scrambling. *Brain and Language*, 86(2), 243–271. [http://doi.org/10.1016/S0093-934X\(02\)00543-6](http://doi.org/10.1016/S0093-934X(02)00543-6)
- van Herten, M., Chwilla, D. J., & Kolk, H. H. J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, 18(7), 1181–97. <http://doi.org/10.1162/jocn.2006.18.7.1181>
- van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22(2), 241–55. <http://doi.org/10.1016/j.cogbrainres.2004.09.002>
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, 18(4), 380–

393.

- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2), 176–190.
<http://doi.org/10.1016/j.ijpsycho.2011.09.015>
- Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *Journal of Experimental Child Psychology*, 99, 75–95. <http://doi.org/10.1016/j.jecp.2007.10.004>
- Williams, A. (2015). *Arguments in syntax and semantics*. Cambridge, UK: Cambridge University Press.