

UC Davis

UC Davis Electronic Theses and Dissertations

Title

Examining The Competing Interpretations Hypothesis in the Comprehension of Noncanonical Structures

Permalink

<https://escholarship.org/uc/item/6v31r045>

Author

Chantavarin, Suphasiree

Publication Date

2021

Peer reviewed|Thesis/dissertation

Examining The Competing Interpretations Hypothesis in the Comprehension of Noncanonical Structures

By

SUPHASIREE CHANTAVARIN
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Fernanda Ferreira, Chair

Tamara Y. Swaab

David P. Corina

Committee in Charge

2021

ABSTRACT

Most current models of syntactic parsing assume that the language comprehension system computes detailed, veridical representations of the linguistic input. Contrary to this assumption, numerous studies have found that people systematically misinterpret unambiguous sentences that require a noncanonical order of thematic role assignment as well as sentences that require the construction of elided linguistic content, suggesting that the parser may perform a shallow, semantic-based analysis in addition to a detailed compositional analysis. However, it remains unclear how the syntactic and semantic streams are reconciled to generate a final interpretation that is sometimes incompatible with the veridical interpretation. The goal of this dissertation was to assess the hypothesis that the misinterpretation of noncanonical sentences reflects a competition between the veridical interpretation and an alternative, highly plausible interpretation that are considered in parallel (i.e., the competing interpretations hypothesis).

Chapter 1 provides a brief overview of previous research and theoretical perspectives on the comprehension of noncanonical structures and describes the motivations of the competing interpretations hypothesis. Chapter 2 describes two experiments that assessed the competing interpretations hypothesis in the comprehension of passive sentences. Results showed that misinterpretation rates were higher as the plausibility of the alternative interpretation increased, but this effect was observed only when the alternative interpretation was more plausible than the veridical interpretation, and not when the two interpretations were equally plausible. Chapter 3 outlines the rationale for a follow-up study that extends these findings by investigating the competing interpretations hypothesis in ellipses, another type of movement-derived structure that

may also be subject to misinterpretations. Taken together, these experiments provide evidence against the competing interpretations hypothesis that assumed a parallel, competitive architecture in sentence processing, and instead lends support to two-stage models of parsing that assume an initial commitment to one interpretation that may be subsequently revised when a syntactic or semantic anomaly is detected.

ACKNOWLEDGEMENTS

First and foremost, I owe my deepest gratitude to my mentor Dr. Fernanda Ferreira for taking me under her wing and guiding me through the wide expanse of psycholinguistic research, and for pushing me farther than I thought I could go. Her unwavering support, encouragement, compassion, and expertise have been instrumental in my growth as a cognitive scientist. I also would like to thank Dr. Tamara Swaab, Dr. David Corina, Dr. Debra Long, and Dr. Raul Aranovich, among others, for their valuable feedback on this project.

My appreciation extends to all my colleagues and friends at Davis, especially the Ferreira Lab current members and alumni for being a wellspring of moral support and inspiration, and for always being willing to share knowledge and constructive advice throughout the years. Special thanks go to the research assistants who helped me on various parts of this project; to Hossein Karimi for all the help I received in my first year as a graduate student; and to Harvey Qiu for the ample encouragement and support both as a friend and as an insightful colleague. To my dearest friends and fellow scientists Nora Beier and Zoe Yang, our tight friendship has made this endeavor infinitely more fun and heartwarming than I could have ever imagined. I am grateful and honored that the three of us got to do this academic journey together.

I would like to acknowledge the Psychology department at Chulalongkorn University (CU) in Thailand for granting me the opportunity and the financial support to pursue a PhD degree in the US. I am grateful to the faculty members and staff in the department at CU who have supported me in this endeavor from start to finish.

I also owe a debt of gratitude to my beloved partner, Frank. Your daily words of love, encouragement, and support has been the bedrock that carried me through all the stress and challenges I faced. Last but not least, I am fully indebted to my family. To my dearest Dad and Mom, my brother P'Nunu and my sister-in-law P'Beau, thank you for your enduring love and support in helping me fulfill my ambition. This accomplishment wouldn't be possible without you.

TABLE OF CONTENTS

Chapter 1: Introduction	1
References	13
Chapter 2: Does the misinterpretation of passive sentences reflect competing interpretations?	21
Abstract	21
Introduction	22
Experiment 1	32
Participants	33
Materials	33
Procedure	39
Analysis	40
Results: A Priori Analyses	43
Results: Exploratory Analyses	51
Discussion	56
Experiment 2	58
Participants	59
Materials	59
Procedure	60
Analysis	60
Results: A Priori Analyses	62
Results: Exploratory Analyses	69
Comparing Experiments 1 & 2	71
General Discussion	74
References	85
Appendix	91
Chapter 3: Shallow processing in the comprehension of elliptical verb phrases	95
Abstract	95
Introduction	96

Absolute plausibility norming study	109
Relative plausibility norming study	114
Data collection plans	117
Data analysis plans	118
References	121
Appendix	126
Chapter 4: Conclusions	132
References	135

CHAPTER 1

Introduction

The main goal of language comprehension is to recover the message that the speaker/writer has conveyed through the linguistic input. Of particular interest is how people understand events that are conveyed via language. Communicating about events is integral to everyday language use, such as when telling someone about an experience, reading a novel, or listening to a podcast. The comprehension of linguistic events crucially relies on the understanding of who did what to whom in a sentence (thematic role assignment), a cognitive process that establishes the thematic relationship between different entities in a sentence and identifies who did what to whom in the discourse. Thematic role assignment is a relatively straightforward process in simple structures such as the active voice that mentions the role higher on a thematic hierarchy before lower roles (e.g., the agent or experiencer before the patient or theme), which is the canonical order in English and in many other languages (Kemmerer, 2012). In noncanonical structures such as passive-voice sentences, however, the process of thematic role assignment can lead to a processing cost in terms of reduced comprehension accuracy (Street & Dabrowska, 2006; Jackson, Lorimor, & Van Hell, 2020; Amichetti, White, & Wingfield, 2016). More broadly, prior work on the processing of noncanonical structures has shown that comprehenders sometimes assign a final interpretation that is inconsistent with the linguistic input, or so-called misinterpretation effects in comprehension (Ferreira, 2003; Slattery et al., 2013; Gibson, Bergen, & Piantadosi, 2013; Christianson, 2016). This dissertation examines the factors that affect thematic role assignment in two types of noncanonical structures: passive sentences and ellipses. Specifically, the experiments described in this dissertation test the hypothesis that the

misinterpretation of noncanonical structures can be explained by the availability of an alternative, highly plausible interpretation that competes with the veridical interpretation of the sentence.

Because passive sentences mention the patient before the agent, the thematic roles have to be assigned in a noncanonical order. For example, in the passive sentence *The thief_{PATIENT} was arrested by the police_{AGENT}*, the preverbal argument *thief* is the patient of the verb *arrest*, whereas the postverbal argument *police* is the agent of the action. The relative difficulty of processing passives has garnered much interest in various strands of research, including in healthy adults (Paolazzi, Grillo, Alexiadou, & Santi, 2019; Street & Dabrowska, 2014; Lim & Christianson, 2013), clinical populations (Friederici & Graetz, 1987; Durrleman, Delage, Prévost, & Tuller, 2017; Aziz, Hassan, Razak, & Garraffa, 2020), and in children acquiring their first language (Abbot-Smith, Chang, Rowland, Ferguson, & Pine, 2017; Huang, Zheng, Meng, & Snedeker, 2013; Messenger, Branigan, McLean, & Sorace, 2012). Another noncanonical structure considered in this dissertation is ellipses, a type of anaphora in which a portion of the sentence is elided or omitted, and thus the interpretation of the elliptical clause depends on the exact form and meaning of the preceding text (i.e., the antecedent; see Phillips & Parker, 2014 for a review on the processing of ellipses). The following sentence pair illustrates the anaphoric nature of ellipsis: *Jane had gone to work. Mark had too*. To interpret the second sentence, the language comprehension system must retrieve a syntactically matching antecedent from the previous sentence (“...had gone to work”) and use this information to assign the linguistically correct interpretation to the second sentence (*Mark had gone to work too*). The remaining sections of this Introduction set the stage for the ensuing chapters by providing a brief overview of the prominent theoretical accounts and debates that are relevant to the discussion of misinterpretation effects in noncanonical structures, followed by a

description of the competing interpretations hypothesis that is assessed in the experiments reported in this dissertation.

1.1 Views on syntactic parsing: then and now

Thematic role assignment is a component process of syntactic parsing, which establishes the dependencies between the words in a sentence according to language-specific syntactic rules and assigns an interpretation to the linguistic input. Various cues are available to the language comprehension system in the service of parsing, including syntactic, semantic, and prosodic cues, as well as the broader discourse context. A fundamental goal in psycholinguistic research is to characterize how these different cues inform parsing, as well as the time course at which they exert influence. The extent to which semantic and discourse information influence parsing decisions in the face of lexical ambiguity and syntactic ambiguity has been extensively debated (e.g., Ferreira & Clifton, 1986; Altmann & Steedman, 1988; Binder, Duffy, & Rayner, 2001; Trueswell, Tanenhaus, & Garnsey, 1994).

Various theoretical accounts of syntactic parsing have been put forward to characterize how semantic and syntactic cues are integrated during incremental processing. Two widely studied models of parsing that were proposed a few decades ago are the syntax-first or garden-path model (Frazier & Fodor, 1978; Frazier, 1979; Rayner, Carlson, & Frazier, 1983) and the constraint-based lexicalist models (MacDonald, Pearlmutter, & Seidenberg, 1994; see also McRae, Spivey-Knowlton, & Tanenhaus, 1998; Spivey-Knowlton & Sedivy, 1995). Crucially, these two types of models differ in their assumptions regarding the architecture and dynamics of the syntactic processor. On the one hand, the two-stage model assumes that parsing is a serial, bottom-up process that deterministically constrains the initial syntactic and semantic interpretation of the

sentence based on certain parsing heuristics (Fodor & Frazier, 1980; also see Ferreira & Clifton, 1986). If a syntactic or semantic anomaly is detected later, reanalysis will take place to repair the initially constructed analysis (Frazier & Clifton, 1998; Fodor & Inoue, 1994; Rayner, Carlson, & Frazier, 1983). On the other hand, constraint-based models assume that the parser operates in a parallel and interactive manner, in which all the available cues are free to dynamically interact to determine the most likely interpretation based on various constraints such as frequency of occurrence and compatibility with the context (MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell, Tanenhaus, & Garnsey, 1994). This divergence in the assumed architecture and dynamics of the parser will be revisited later in the Introduction. Despite these differences, both of these older models of parsing share the assumption that the syntactic cues are absolute, such that the ultimate interpretation is always consistent with the syntactic structure.

Contrary to this latter assumption, subsequent experiments have shown that people sometimes assign a final interpretation that is incompatible with the detailed syntactic analysis of the linguistic input. This has been extensively shown for garden-path sentences (Christianson, 2008; Christianson, Williams, Zacks, & Ferreira, 2006; Ferreira, Christianson, & Hollingworth, 2001), and the initial, wrong interpretation may remain incomplete even if syntactic reanalysis takes place to resolve the temporary ambiguity (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Slattery et al., 2013; Sturt, 2007). Beyond garden-path sentences, there is evidence that even unambiguous, noncanonical sentences can be systematically misinterpreted. A prime example is the passive construction, which has been shown to be misinterpreted especially when the sentence meaning is implausible (Ferreira, 2003). Similarly, implausible elliptical clauses have also been

shown to be misinterpreted when there is another meaning suggested by the context that is more plausible than the veridical interpretation (Garnham & Oakhill, 1987).

These findings indicate that the final interpretation may not always be faithful to the linguistic input, which is not predicted by the aforementioned models of parsing. More recent models of parsing have attempted to explain these misinterpretation effects, such as the good-enough model (Ferreira & Patson, 2007; Ferreira, Bailey, & Ferraro, 2002) and the noisy channel model of parsing (Gibson, Bergen, & Piantadosi, 2013). The good-enough approach to language comprehension attributes misinterpretations to shallow processing. Under this view, misinterpretations occur because people may not always construct detailed, complete, and veridical representations of the input. A related account is the noisy channel model, which takes a rational, probabilistic approach and proposes that misinterpretations reflect the parser's attempt to correct an implausible sentence into a more plausible one. Various factors are considered in this correction process, including perceptual and production errors, as well as other sources of noise in the environment under which communication occurs.

Misinterpretation effects are an important topic to investigate because they can inform us about how phrase structure rules on the one hand, and the meaning of words and expressions on the other hand, are integrated in real time to generate the final sentence interpretation. The evidence for misinterpretations from offline comprehension measures dovetails with the more general findings that plausibility and contextual information exert a strong influence during online language processing, particularly when the sentence contains conflicting syntactic and semantic cues. Event-related potential (ERP) studies on sentences containing semantic reversal anomalies (e.g., *The fox that hunted the poachers...*) have found that the critical words (*poachers* in the preceding

example) do not elicit an N400 effect but rather a late posterior positivity (i.e., the P600 effect; Van Herten, Chwilla, & Kolk, 2006; Van Herten, Kolk, & Chwilla, 2005; Zheng & Lemhöfer, 2019), despite the fact that the N400 component is typically sensitive to the ease of semantic activation and retrieval (Kuperberg, 2016; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Kuperberg, Brothers, & Wlotko, 2020) whereas the posterior P600 has been associated with syntactic processing and revision (Osterhout & Nicol, 1999; Kaan, Harris, Gibson, & Holcomb, 2000; Fiebach, Schlesewsky, & Friederici, 2002). This so-called semantic P600 effect may reflect thematic processing difficulty (Hoeks, Stowe, & Doedens, 2004), and/or a conflict between the representations constructed in parallel by syntactic parsing and semantic or plausibility-based processing (Nakano, Saron, & Swaab, 2010; Kim & Osterhout, 2005; Kolk et al., 2003, Van Herten et al., 2006, Van Herten et al., 2005; for critical reviews see Bornkessel-Schlesewsky & Schlesewsky, 2008, Brouwer, Fitz, & Hoeks, 2012). The latter explanation is formalized in the dual-route account of language comprehension, which proposes that parsing may entail algorithmic, syntactic processing as well as a potentially independent stream of semantic processing (Kuperberg, 2007; Dwivedi, 2013; Karimi & Ferreira, 2016; Van Herten, Kolk, & Chwilla, 2005; Kim & Osterhout, 2005; Hagoort, Baggio, & Willems, 2009). This echoes Ferreira's (2003) proposal that the language comprehension system may perform not only a detailed syntactic analysis but also a heuristics-based analysis (e.g., relying on the agent-first heuristic and the plausibility heuristic). Note that the proposal of two separate routes to comprehension has been subsequently challenged by models that instead appeal to other mechanisms such as probabilistic inference (Kuperberg, 2016; Kuperberg, Brothers, & Wlotko, 2020) and cycles of semantic

retrieval and integration (Brouwer, Crocker, Venhuizen, & Hoeks, 2017; Brouwer, Fitz, & Hoeks, 2012), without the need to invoke multiple processing streams.

1.2 *Competing interpretations hypothesis*

Whether one adopts the position that there are two complementary streams of processing or not, an outstanding issue in the processing of noncanonical structures is to identify a specific cognitive mechanism that can account for the observed misinterpretation effects in the final interpretation. From the dual-route perspective, if there are potentially two streams to comprehension, how are the output from the syntactic stream and the semantic stream reconciled? Conversely, in terms of constraint satisfaction and rational inference, how are the syntactic and semantic cues weighted to derive a final interpretation that is sometimes not the veridical one? To address this theoretical gap, this dissertation investigates the influence of event-based plausibility information on the final interpretation of passive sentences (Chapter 2) as well as passive elliptical verb phrases (Chapter 3). Specifically, we propose the *competing interpretations hypothesis* as a potential explanation for the misinterpretation of noncanonical structures and test this hypothesis in the experiments described in the ensuing chapters. The competing interpretations hypothesis proposes that the availability of a highly plausible, alternative interpretation increases the likelihood of misinterpretations. This hypothesis is motivated by models of parsing which assume that multiple analyses can be considered at once (e.g., Jackendoff, 2007; MacDonald, Pearlmutter, & Seidenberg, 1994; Tabor, Juliano, & Tanenhaus, 1997). Most of these models further assume that the different possible analyses under consideration are bounded and ranked, in which the parser narrows down to a small set of analyses that are ordered according to how much they are supported by the various cues in the sentence. Under these parallel-processing models, parsing is

also assumed to be highly interactive, in which the semantic and discourse context, including plausibility information, can immediately influence incremental parsing decisions. Crucial to our hypothesis, a subset of these models such as the Processing Competition model (Hagoort, Baggio, & Willems, 2009; Kos, Vosse, Van Den Brink, & Hagoort, 2010) further propose that there may be competition between the highly ranked analyses (for critical reviews of the parallel processing architecture, see Clifton & Staub, 2008 and Vosse & Kempen, 2009). However, unlike the original accounts of parallel-competitive parsing, the competing interpretations hypothesis further assumes that competition between multiple analyses may occur even when one of the activated meanings is clearly not supported by the syntactic cues.

To account for the misinterpretation of noncanonical structures, we assume a combination of the parallel-competitive architecture along with the idea of shallow processing as proposed by the good-enough model of parsing. During incremental thematic processing of passive sentences, at the onset of the by-phrase (“*was verb-ed by...*”) the parser assigns the preverbal argument as the patient or recipient of the action and expects an upcoming agent in the sentence (Liversedge, Pickering, Branigan, & Van Gompel, 1998). Ultimately, the parser will generate the veridical meaning of the sentence according to the syntactic cues (this process may be fully specified or underspecified; see Traxler, 2014). Crucially, if the parser does rely on both syntactic and semantic streams of processing, the semantic stream will also deliver an interpretation. According to the competing interpretations hypothesis, if multiple analyses are considered in parallel, we might expect some competition and therefore interference between the veridical, syntax-driven interpretation and non-veridical, semantics-based interpretation.

In the experiments reported in this dissertation, we test the competing interpretations hypothesis in two main types of passive sentences that differ in the degree of meaning bias with respect to the order of the verbal arguments, illustrated in examples (1), (2), and (3) below. The first type are passive sentences that contain a syntax-semantics conflict, such as in (1b) and (2b), and the second type are passives that have two equally plausible meanings, such as in (3). In (1), the two noun phrases are nonreversible because one argument is inanimate, and their reversal would result in an anomalous interpretation (1b). In (2), the order of the noun phrases is biased towards one interpretation (2a) over the other (2b). Finally, in (3), the two argument orders yield equally plausible meanings.

(1) a. The apron was worn by the chef.

b. The chef was worn by the apron.

(2) a. The man was bitten by the dog.

b. The dog was bitten by the man.

(3) a. The customer was thanked by the clerk.

b. The clerk was thanked by the customer.

In the case of syntax-semantics conflict in which the passive sentence's veridical interpretation is implausible such as in examples (1b) and (2b) above, the alternative, more plausible interpretation may be ranked higher (or, given a higher weight) and compete for selection. There is some evidence that this competition occurs during the online processing of passive sentences that contain a syntax-semantics conflict (Thothathiri, Asaro, Hsu, & Novick, 2018). Here, we further propose that the plausibility of this role-reversed interpretation is directly proportional to the likelihood that it will be chosen as the final interpretation. The more plausible the alternative interpretation, the more likely it will be selected. In the case where the passive sentence delivers

two equally plausible meanings such as in example (3) above, we might also expect some competition. In these symmetrically-plausible passive sentences, one meaning is not more likely than the other, and the veridical meaning is perfectly plausible. Yet, the availability of an alternative, equally compelling interpretation may result in misinterpretations. According to constraint-based and competition-based views of parsing, both the veridical and the role-reversed interpretation in these symmetrical passives would be highly preferred and are both highly compatible with the context, and therefore we might expect that there will be strong competition between these two alternatives in the race for the final interpretation. Additionally, either interpretation may be “good-enough” to generate a fairly accurate discourse model (e.g., whether customers thanking clerks or vice versa), since both interpretations support the denoted event in these symmetrical sentences that often involve a reciprocal action (e.g., thanking, greeting, or hugging). Thus, under the view of good-enough processing and parallel-competitive parsing, we might expect competition from an alternative, highly plausible meaning even when that meaning is equally plausible to the linguistically correct interpretation. We note that this is a tentative prediction because the shallow processing of symmetrically plausible sentences is relatively less understood, compared to the shallow processing of sentences containing a syntax-semantics conflict.

These predictions under the competing interpretations can be contrasted with the predictions based on serial models of parsing, which assume that only one analysis is considered at a given time (see Table 1 for a summary of the predictions). Most serial models of parsing are modular, giving precedence to syntactic cues at least in the initial analysis. Under this view, there is no “competition” in the sense that more than one interpretation can actively compete for

selection. Nonetheless, serial models do allow for feedback from the subsequent thematic processor. If the initially selected interpretation has a syntactic or semantic anomaly, the parser will engage in a structural and/or semantic revision. This tendency has been termed “revision as a last resort”, which states that the parser will revise the initial interpretation only if it contains an anomaly or ungrammaticality (Fodor & Inoue, 1994, 2000). Serial models therefore predict that passive sentences containing a syntax-semantics conflict will trigger a semantic revision, but these models do not predict that passive sentences with two equally plausible interpretations will require revision. In this latter case, the parser should commit to the linguistically correct interpretation because it does not convey any anomaly. Given these differences between parallel-competitive and serial models of parsing, contrasting predictions were derived for the symmetrical passive sentences in particular.

In sum, this dissertation reports a series of experiments to assess the competing interpretations hypothesis in accounting for the misinterpretation effects in noncanonical structures. Chapter 2 reports two experiments assessing this hypothesis in the comprehension of passive sentences. Chapter 3 extends this investigation to the comprehension of ellipses. Chapter 4 summarizes the general conclusions from these studies.

Table 1

Predictions based on contrasting hypotheses

Sentence type	Hypothesis	
	Competing interpretations hypothesis (assumes parallel activation)	Alternative hypothesis (e.g., serial thematic role revision)
Nonreversible (e.g., <i>The chef was worn by the apron</i>)	Comprehension accuracy will be lower as the plausibility of the alternative, role-reversed interpretation increases	Comprehension accuracy will be lower as the plausibility of the alternative, role-reversed interpretation increases
Biased (e.g., <i>The dog was bitten by the man</i>)	Comprehension accuracy will be lower as the plausibility of the alternative, role-reversed interpretation increases	Comprehension accuracy will be lower as the plausibility of the alternative, role-reversed interpretation increases
Symmetrical (e.g., <i>The customer was thanked by the clerk</i>)	Comprehension accuracy will be lower as the plausibility of the alternative, role-reversed interpretation increases	NO relationship between comprehension accuracy and the plausibility of the alternative, role-reversed interpretation

References

- Abbot-Smith, K., Chang, F., Rowland, C., Ferguson, H., & Pine, J. (2017). Do two and three year old children use an incremental first-NP-as-agent bias to process active transitive and passive sentences?: A permutation analysis. *PLoS One*, *12*(10), e0186129.
- Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*(3), 191-238.
- Amichetti, N. M., White, A. G., & Wingfield, A. (2016). Multiple solutions to the same problem: utilization of plausibility and syntax in sentence comprehension by older adults with impaired hearing. *Frontiers in Psychology*, *7*, 789.
- Aziz, M. A. A., Hassan, M., Razak, R. A., & Garraffa, M. (2020). Syntactic abilities in Malay adult speakers with aphasia: a study on passive sentences and argument structures. *Aphasiology*, *34*(7), 886-904.
- Binder, K. S., Duffy, S. A., & Rayner, K. (2001). The effects of thematic fit and discourse context on syntactic ambiguity resolution. *Journal of Memory and Language*, *44*(2), 297-324.
- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*, *59*(1), 55-73.
- Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, *41*, 1318-1352.
- Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127-143.
- Christianson, K. (2008). Sensitivity to syntactic changes in garden path sentences. *Journal of Psycholinguistic Research*, *37*, 391-403.

- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368-407.
- Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes*, *42*, 205-238.
- Clifton Jr, C., & Staub, A. (2008). Parallelism and competition in syntactic ambiguity resolution. *Language and Linguistics Compass*, *2*(2), 234-250.
- Dąbrowska, E., & Street, J. (2006). Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences*, *28*(6), 604-615.
- Durrleman, S., Delage, H., Prévost, P., & Tuller, L. (2017). The comprehension of passives in autism spectrum disorder. *Glossa*, *2*(1), 88.
- Dwivedi, V. D. (2013). Interpreting quantifier scope ambiguity: Evidence of heuristic first, algorithmic second processing. *PLoS One*, *8*(11), e81461.
- Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75-84.
- Ferreira, F., & Clifton Jr, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348-368.
- Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, *1*(1-2), 71-83.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164-203.

- Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11-15.
- Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research*, 30, 3-20.
- Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, 47(2), 250-272.
- Fodor, J. D., & Frazier, L. (1980). Is the human sentence parsing mechanism in ATN? *Cognition*, 8, 417-459.
- Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research*, 23(5), 407-434.
- Fodor, J. D., & Inoue, A. (2000). Garden path re-analysis: Attach (anyway) and revision as last resort. In *Cross-linguistic perspectives on language processing* (pp. 21-61). Springer, Dordrecht.
- Frazier, L., & Clifton, C. (1998). Sentence reanalysis, and visibility. In *Reanalysis in sentence processing* (pp. 143-176). Springer, Dordrecht.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6(4), 291-325.
- Frazier, L. (1979). On comprehending sentences: Syntactic parsing strategies [Unpublished doctoral dissertation]. Indiana University Linguistic Club, Bloomington, IN.
- Friederici, A. D., & Graetz, P. A. (1987). Processing passive sentences in aphasia: Deficits and strategies. *Brain and Language*, 30(1), 93-105.

- Garnham, A., & Oakhill, J. (1987). Interpreting elliptical verb phrases. *The Quarterly Journal of Experimental Psychology Section A*, 39(4), 611-627.
- Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, 110(20), 8051-8056.
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15(6), 883-899.
- Hagoort, P., Baggio, G., and Willems, R. (2009). "Semantic unification," in *The New Cognitive Neuroscience*, ed. M. S. Gazzaniga (Cambridge, MA: MIT Press), 56, 1-18.
- Hoeks, J. C., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19(1), 59-73.
- 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.
- Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Research*, 1146, 2-22.
- Jackson, C. N., Lorimor, H., & van Hell, J. G. (2020). The ERP correlates of thematic role assignment for passive versus active sentences. *Journal of Neurolinguistics*, 54, 100886.
- Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15(2), 159-201.

- Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, 69(5), 1013-1040.
- Kemmerer, D. (2012). The cross-linguistic prevalence of SOV and SVO word orders reflects the sequential and hierarchical representation of action in Broca's area. *Language and Linguistics Compass*, 6(1), 50-66.
- 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.
- Kos, M., Vosse, T. G., Van Den Brink, D., & Hagoort, P. (2010). About edible restaurants: conflicts between syntax and semantics as revealed by ERPs. *Frontiers in Psychology*, 1, 222.
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23-49.
- Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*, 31(5), 602-616.
- Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, 32(1), 12-35.
- Lim, J. H., & Christianson, K. (2013). Second language sentence processing in reading for comprehension and translation. *Bilingualism*, 16(3), 518.
- Liversedge, S. P., Pickering, M. J., Branigan, H. P., & van Gompel, R. P. (1998). Processing arguments and adjuncts in isolation and context: The case of by-phrase ambiguities in passives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(2), 461.

- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38(3), 283-312.
- Messenger, K., Branigan, H. P., McLean, J. F., & Sorace, A. (2012). Is young children's passive syntax semantically constrained? Evidence from syntactic priming. *Journal of Memory and Language*, 66(4), 568-587.
- Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience*, 22(12), 2886-2898.
- Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes*, 14(3), 283-317.
- Paolazzi, C. L., Grillo, N., Alexiadou, A., & Santi, A. (2019). Passives are not hard to interpret but hard to remember: evidence from online and offline studies. *Language, Cognition and Neuroscience*, 34(8), 991-1015.
- Phillips, C., & Parker, D. (2014). The psycholinguistics of ellipsis. *Lingua*, 151, 78-95.
- Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22(3), 358-374.

- Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Linger-
ing misinterpretations of garden path sentences arise from competing syntactic
representations. *Journal of Memory and Language*, 69(2), 104-120.
- Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple
constraints. *Cognition*, 55(3), 227-267.
- Street, J., & Dabrowska, E. (2014). Lexically specific knowledge and individual differences in adult
native speakers' processing of the English passive. *Applied Psycholinguistics*, 35(1), 97-118.
- Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, 105(2), 477-488.
- Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-
based account of the interaction of lexical and structural constraints in sentence
processing. *Language and Cognitive Processes*, 12(2-3), 211-271.
- Thothathiri, M., Asaro, C. T., Hsu, N. S., & Novick, J. M. (2018). Who did what? A causal role for
cognitive control in thematic role assignment during sentence
comprehension. *Cognition*, 178, 162-177.
- Traxler, M. J. (2014). Trends in syntactic parsing: Anticipation, Bayesian estimation, and good-
enough parsing. *Trends in Cognitive Sciences*, 18(11), 605-611.
- 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*, 33(3), 285-318.
- Van Herten, M., Chwilla, D. J., & Kolk, H. H. (2006). When heuristics clash with parsing
routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive
Neuroscience*, 18(7), 1181-1197.

- Van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22(2), 241-255.
- Vosse, T., & Kempen, G. (2009). In defense of competition during syntactic ambiguity resolution. *Journal of Psycholinguistic Research*, 38(1), 1.
- Zheng, X., & Lemhöfer, K. (2019). The “semantic P600” in second language processing: When syntax conflicts with semantics. *Neuropsychologia*, 127, 131-147.

CHAPTER 2

Does the misinterpretation of passive sentences reflect competing interpretations?

Abstract

Previous research has shown that comprehenders systematically misinterpret passive sentences whose meaning contradicts with schematic knowledge about event plausibility. However, it is unclear how the syntactic and semantic cues in the sentence are reconciled to generate a final interpretation that is sometimes inconsistent with the veridical interpretation. In this study, we investigated the hypothesis that the misinterpretation of passive sentences is driven by a concurrent competition between the veridical interpretation and an alternative, role-reversed interpretation that is strongly supported by schematic knowledge. In a reading comprehension study (Experiment 1) and a listening comprehension study (Experiment 2), participants were presented with sentences that varied in structure (active or passive) and in the order of the verbal arguments (plausible/order1 or implausible/order2). In both experiments, accuracy on the comprehension questions was predicted by ratings of the role-reversed interpretation's plausibility when the sentences denoted a syntax-semantics conflict (e.g., *The secretary was typed by the letter; The dog was bitten by the man*), but this effect was absent when the sentences contained two symmetrically plausible interpretations (e.g., *The customer was thanked by the clerk*). Taken together, these results provide evidence against the competing interpretations hypothesis that assumed parallel-competitive activation of multiple analyses, and instead the findings support two-stage models of parsing that assume an initial commitment to a single syntactic analysis which may be subject to subsequent thematic role revision when a semantic anomaly is detected.

Keywords: sentence comprehension, thematic role assignment, misinterpretation errors, passives, event plausibility, competition

1. Introduction

To interpret the meaning conveyed by a sentence, the language comprehension system must reconcile information from the sentence's syntactic structure as well as lexical-semantic information, which together determine the assignment of thematic roles denoting who did what to whom in the sentence. For instance, in the simple active-voice sentence *The girl kicked the ball*, the *girl* is the entity doing the action of kicking (i.e., the agent) whereas the *ball* is the recipient of that action (i.e., the patient). When processing noncanonical sentences, however, comprehenders sometimes assign a semantic interpretation that is inconsistent with the syntactic frame, according to recent models of parsing (e.g., Gibson, Bergen, & Piantadosi, 2013; Ferreira & Patson, 2007; Ferreira, Bailey, & Ferraro, 2002). Such misinterpretation effects have been shown for not only temporarily ambiguous sentences (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Sanford, 2002), but also syntactically unambiguous sentences. For instance, people tend to systematically misinterpret implausible passive sentences (e.g., *the dog was bitten by the man*) by selecting a lexically-driven interpretation that is more plausible (*the dog bit the man*; Ferreira, 2003; Christianson et al., 2010; Gibson, Bergen, & Piantadosi, 2013; Dabrowska & Street, 2006; Amichetti, White, & Wingfield, 2016). Given the findings that people may not always have a veridical internal representation of unambiguous, noncanonical sentences such as passive sentences, it is important to elucidate the mechanisms that can explain this misinterpretation effect. However, the exact mechanisms underlying such misinterpretations are not well-understood. The current study investigates whether the misinterpretation of passive sentences can be explained by the extent to which an alternative, role-reversed interpretation competes with the sentence's veridical interpretation (i.e., the *competing interpretations hypothesis*).

1.1 Shallow processing in noncanonical structures

One model of language comprehension that is relevant to the misinterpretation of passive sentences is the good-enough model of parsing (Ferreira & Patson, 2007; Ferreira, Bailey, & Ferraro, 2002). According to this model, linguistic representations may not always be complete nor accurate because comprehension is achieved in a “good-enough” manner for the communicative purpose at hand. In some instances, the parser may not engage in detailed, structural analysis of the sentence, but rather shallow processing may take place (for further discussion, see Karimi & Ferreira, 2016). Incomplete interpretations may persist even if syntactic reanalysis takes place, and this claim is supported by evidence from studies showing lingering effects in the processing of garden-path sentences (Sturt, 2007; Slattery et al., 2013; Christianson, Williams, Zacks, & Ferreira, 2006). When processing a temporarily ambiguous sentence such as *While Mary bathed the baby played in the crib*, people have to revise the syntactic structure when encountering the verb *played* if they realize that Mary did not bathe the baby, but rather that Mary bathed herself. Evidence from comprehension questions on these garden-path sentences suggest that people try to resolve the temporary structural ambiguity, but comprehenders may nevertheless come away with an incomplete interpretation in which *the baby* is both the subject of *played* and also the object of *dressed*. These findings suggest that the initially constructed interpretation may linger and compete with the veridical interpretation, even after syntactic reanalysis has taken place.

Besides garden-path sentences, unambiguous, noncanonical sentences like passive sentences may also be processed in a shallow manner. Comprehenders may rely on heuristics about thematic ordering (e.g., in SVO languages such as English, a canonical sentence typically mentions the agent before the patient) and on plausibility heuristics (e.g., language is typically used

to convey plausible utterances), which can lead to an incorrect interpretation of the sentence. A foundational study by Ferreira (2003) provided evidence for the shallow processing of passive sentences. In the study, participants listened to syntactically unambiguous sentences (actives, passives, subject-clefts, and object-clefts), and after each sentence they were asked to name the agent or the patient of the described event. The accuracy data showed that people systematically misinterpreted sentences that required thematic roles to be assigned in a noncanonical order (e.g., the patient before the agent), such as *the dog was bitten by the man* and *it was the man the dog bit*. Misinterpretations were also more likely when the order of the arguments resulted in an implausible or anomalous meaning, such as *the man bit the dog* compared to the plausible version *the dog bit the man*. The likelihood of misinterpretation was highest for the passive implausible sentences in particular, which violated real-world plausibility as well as the typical order of thematic role assignment. Based on these findings, Ferreira (2003) concluded that sentence interpretation is derived by not only algorithmic syntactic processing but also heuristics such as preference for linearly assigning the agent before the patient, and preference for an interpretation that conforms to schematic knowledge about events (for an alternative view, see Bader & Meng, 2018 and Meng & Bader, 2021). Numerous other studies have found similar results supporting the idea of shallow processing in noncanonical sentences (Christianson et al., 2010; Gibson, Bergen, & Piantadosi, 2013; Dabrowska & Street, 2006; Amichetti, White, & Wingfield, 2016).

Event-related potential (ERP) studies on the processing of sentences that contain thematic role violations provide further evidence that comprehenders compute a semantics-based interpretation in addition to the compositional interpretation. In these studies, subjects are presented with semantically implausible sentences such as “*For breakfast the eggs would only eat...*”

(Kuperberg, Sitnikova, Caplan, & Holcomb, 2003) or “*The hearty meal was devouring the kids*” (Kim & Osterhout, 2005), and participants’ neural response to the critical word (underlined) is measured. The anomaly in these sentences is that the thematic roles of the verb are contrary to what is expected based on semantic and pragmatic information. These sentences with thematic role violations elicited a smaller N400 effect compared to sentences with non-thematic role semantic violations, and surprisingly, they also elicited a posterior P600 effect. One interpretation of this so-called semantic P600 effect to thematic role violations is that it reflects an online clash between the semantic and syntactic streams in language processing when the output from these two streams are in conflict (Kuperberg, 2007; Kolk, Chwilla, Van Herten, & Oor, 2003; Van Herten, Chwilla, & Kolk, 2006; Van Herten, Kolk, & Chwilla, 2005; for an alternative view, see Bornkessel-Schlesewsky & Schlewsky, 2008). Failure to resolve this syntax-semantics conflict may lead to significant misinterpretations when participants’ final interpretations are probed, though the conflict may not necessarily impact question answering accuracy if the syntactic and semantic streams were successfully reconciled during online processing (e.g., Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; Nakano, Saron, & Swaab, 2010).

1.2 Competing interpretations hypothesis

The aforementioned studies showed that one of the key factors that influence the likelihood of misinterpretation is event-based knowledge in semantic memory. Event knowledge comprises information about the common actions, participants, and components of an event, such as the typical agents, patients, themes, goals, instruments, and as well as the locations at which the event usually occurs. Crucially, part of event knowledge is knowing who tends to do what to whom in specific situations, such as knowledge that waiters are likely to serve customers, or that doctors

are likely to treat patients, but not the other way around. This type of information has been shown to be computed rapidly from single words (McRae, Ferretti, & Amyote, 1997). Prior research on the role of plausibility strategies in the misinterpretation of passive sentences has addressed how the plausibility of the sentence's veridical meaning affects sentence interpretation (e.g., for the sentence *The dog was bitten by the man*, how plausible is it for people to bite dogs?), and this plausibility variable has typically been used as a dichotomous predictor (e.g., plausible or implausible; e.g., Ferreira, 2003; Christianson et al., 2010; Gibson, Bergen, & Piantadosi, 2013; Dabrowska & Street, 2006; Amichetti, White, & Wingfield, 2016). However, another way to investigate the influence of event knowledge on comprehension is to assess the plausibility of the competing, role-reversed meaning (e.g., for the sentence *The dog was bitten by the man*, how plausible is it for dogs to bite men, and how much more plausible is it for dogs to bite men than the other way around?). This approach allows a more direct investigation of the mechanisms underlying the misinterpretation of passive sentences.

According to the competing interpretations hypothesis that we propose here, the parser may concurrently consider the alternative, role-reversed interpretation of the sentence in addition to the veridical interpretation when assigning an interpretation to noncanonical structures. The competing interpretations hypothesis assumes that the parser can entertain two sets of thematic roles and pick the "better" one, and it predicts that comprehenders will occasionally select the alternative, non-veridical interpretation when that interpretation is compelling or plausible. Under expectation-based and surprisal models of parsing, the parser constructs all possible syntactic analyses using the available syntactic and semantic cues in the input and relies on probabilistic inference as well as animacy cues of the subject noun to select the most likely interpretation

(Nakano, Saron, & Swaab, 2010; Levy, 2008; Gibson, Bergen, & Piantadosi, 2013; Hale, 2003). If more than one interpretation is considered at once, there may be some competition among the highly-ranked interpretations (e.g., Hagoort, Baggio, & Willems, 2009; Kos, Vosse, Van Den Brink, & Hagoort, 2010). Note that parallel versus serial processing is an ongoing major debate in psycholinguistics (for detailed discussions see Lewis, 2000 and Gibson & Pearlmutter, 2000). There are other models of sentence processing that assume serial parsing rather than parallel parsing (e.g., Frazier & Fodor, 1978; Fodor & Inoue, 1998). As will be discussed in the subsequent Current Study section, this distinction between parallel versus serial parsing leads to different predictions regarding the competition that may take place between the veridical interpretation and the non-veridical, role-reversed interpretation in the processing of passive sentences.

Competing interpretations should have an influence on offline comprehension particularly in passive sentences, compared to active sentences. Since the comprehension of passives requires a noncanonical order of thematic role assignment, and given that speakers of English tend to rely on the agent-before-patient strategy, this means that sometimes the thematic roles will have to be revised during the parsing process. In passive constructions, people may have difficulty keeping track of the binding between the thematic roles (agent/patient) and the grammatical roles (subject/oblique), making the thematic roles in these sentences more vulnerable to good-enough processing and semantic-heuristic strategies (for an alternative view on the role of the agent-patient heuristic, see Paolazzi, Grillo, Alexiadou, & Santi, 2019). In passive, implausible sentences such as *The cat was chased by the mouse*, schematic knowledge suggests an alternate set of thematic roles (*cat chased mouse*), and the agent-before-patient strategy also supports this alternative interpretation. Schemas may therefore guide interpretation when the sentence is syntactically complex, such as in

passive sentences that require noncanonical assignment of the patient before the agent. The parser may occasionally select the shallow, schema-consistent but non-veridical interpretation (*cat chased mouse*), rather than using all the available cues in the input to arrive at the veridical interpretation (*mouse chased cat*).

There is some preliminary evidence suggesting that competing interpretations may underlie the processing of passive sentences, although it was not the main research question in these studies. In a conceptual replication of Ferreira (2003), Bader and Meng (2018) found in their exploratory analyses that people were more likely to misinterpret implausible passive sentences (e.g., *The dog was bitten by the man*) when the role-reversed meaning (*The dog bit the man*) was relatively more plausible than the veridical meaning (*The man bit the dog*). Moreover, in a visual world study on the online processing of passive sentences, Thothathiri and colleagues (2018) found that looking times did not differ between the picture denoting the veridical interpretation of passive sentences (e.g., *The rabbit was chased by the fox*) and the picture denoting the role-reversed interpretation (e.g., *The fox was chased by the rabbit*) before the onset of the role disambiguating region of the sentence (“...ed by...”), suggesting that participants considered the preverbal agent *rabbit* as both a possible chaser and flier, and that there may have been concurrent activation and competition between the veridical interpretation and the non-veridical, role-reversed interpretation. Although these experiments suggest the possibility that the processing of passive sentences may entail the activation of competing interpretations, to our knowledge no study has directly investigated whether competing thematic role assignments underlie the misinterpretation of passive sentences.

1.3 Current Study

To address this gap, in this study we assessed whether the plausibility of the non-veridical, role-reversed interpretation increases the likelihood of misinterpreting passive sentences. Novel to the current study, we directly tested the notion of competing interpretations by collecting two continuous measures of event plausibility to assess whether the plausibility of the non-veridical, role-reversed meaning can account for the comprehension of passives. One measure was the absolute plausibility rating of the role-reversed meaning, relative to all other events in the world (henceforth referred to as “role-reversed plausibility”), whereas the other was a novel measure indexing the relative plausibility rating between the veridical interpretation and the role-reversed interpretation (henceforth referred to as “relative plausibility”). In the main experiment, we used the same experimental design and stimuli as in Ferreira’s (2003) study on the comprehension of passive sentences. There were three sentence types based on the two meanings derived from reversing the arguments in the sentence: nonreversible (*The mouse ate the cheese*), biased (*The dog bit the man*), and symmetrical sentences (*The customer thanked the clerk*). In Experiment 1 the sentences were presented in written form, whereas in Experiment 2 they were presented aurally. The experimental task was for subjects to respond to comprehension questions about thematic role assignment after each sentence.

1.3.1 Predictions

When the plausibility of the veridical interpretation (henceforth referred to as the “veridical plausibility” for brevity) is used as the predictor of accuracy and response times on the comprehension questions, we expected to replicate the findings from Ferreira (2003) regarding the processing cost associated with the passive structure and the implausibility of the sentence’s

veridical interpretation. If the noncanonical order of thematic role assignment causes difficulty, we should find lower accuracy and longer response times for passive than active sentences. In addition, if there is a processing cost for sentences that contradict knowledge of event plausibility, accuracy should be lower and response times should be longer for implausible than plausible sentences. We also expect to find lowest accuracy and longest response times for the passive implausible sentences, reflecting the highest degree of processing difficulty for sentences that both require a noncanonical order of thematic role assignment and also contradict event plausibility. Note that the predictions involving veridical plausibility do not apply to the symmetrical items because by design, both meanings are equally plausible.

Importantly, our central hypothesis was that comprehension of passive sentences will be affected not only by the veridical plausibility of the sentence but also by the plausibility of the role-reversed meaning. Under this hypothesis, we expect to find lower accuracy and longer response times when there is a plausible, role-reversed interpretation. This leads to straightforward predictions for the nonreversible and biased sentence types since the role-reversed interpretation is clearly more plausible than the veridical interpretation of the implausible passive sentences. There may be competition between the veridical, implausible meaning and the role-reversed, more plausible meaning that arises from a conflict monitoring process when the language processing system encounters a syntax-semantics conflict, leading to systematic misinterpretations.

For the symmetrical items that were designed to have an equally plausible veridical meaning and role-reversed meaning, there are at least two possible outcomes regarding the influence of the role-reversed interpretation. One possibility is that there is competition between the two sets of thematic roles that are simultaneously and automatically activated, as summarized

in the previous section. This prediction is based on the view of syntactic underspecification which proposes that the parser activates all possible interpretations and selects one that is most plausible or appropriate. If multiple interpretations are automatically activated at once, there are reasons to expect competition between the two equally plausible meanings. Because the symmetrical sentences contain two entities engaging in a reciprocal action (e.g., customers thanking clerks), the two activated sets of thematic roles may be confusable based on similarity-based interference (for detailed discussions of this concept, see Gordon, Hendrick, & Johnson, 2001; Lewis & Vasishth, 2005; Van Dyke & McElree, 2006). For example, the passive sentence *the customer was thanked by the clerk* activates a schema of clerks thanking customers, and vice versa. Both the veridical interpretation and the role-reversed interpretation are supported by the same event schema, and thus either interpretation is a good fit (or “good enough”) to understand the gist of the sentence. A visual world study by Thothathiri and colleagues (2018) further suggests that both the veridical and the role-reversed interpretations are considered in parallel during real-time sentence processing of symmetrically plausible passive sentences. Thus, under this view of automatic, parallel activation, we might find an effect of the competing interpretation for all sentence types, whether the two meanings considerably differ in plausibility (nonreversible and biased types) or when they are equally plausible (symmetrical type).

In contrast to the parallel processing view, serial-type models of parsing assume that only one analysis is assessed at a time (Frazier & Fodor, 1978; Fodor & Inoue, 1998). Under this serial processing view, the veridical interpretation may be revised only if this initially computed meaning is implausible and therefore triggers a thematic reanalysis, as is the case for nonreversible and biased sentences. In contrast, for the symmetrical sentences, there would be no need for

semantic/syntactic revision because the veridical interpretation of the passive symmetrical sentences is already sensible (e.g., whether clerks thanking customers or customers thanking clerks). In sum, under serial views of parsing, there would be no concurrent activation and therefore no competition from the non-veridical, equally plausible meaning. This means we might find an effect of the competing interpretation's plausibility when one of the interpretations is more plausible (e.g., in the nonreversible and biased sentences), but not when both meanings are equally plausible (e.g., in the symmetrical sentences).

2. Experiment 1: reading study

In Experiment 1, we assessed how the plausibility of the role-reversed interpretation influences the comprehension of passive sentences in a reading comprehension study. Participants read sentences that varied in structure (active, passive) and veridical plausibility (plausible/order1, implausible/order2). After each sentence, participants answered a comprehension question asking who was the agent in the denoted event (agenthood was defined to subjects as the “do-er” of the action). This agent decision task was chosen to directly target thematic role mapping within the sentence. Unlike in Ferreira (2003)'s study where participants were required to identify the agent as well as the patient on the experimental stimuli, in the current study participants were probed only about the agent on the experimental items, in order to eliminate interference and confusion caused by having to keep track of the agent-patient distinction in the comprehension task. Instead, questions about the patient were presented on some of the filler items to ensure subjects did not develop a strategy of focusing exclusively on agents. To test our hypothesis regarding the role of competing interpretations in the comprehension of passive sentences, we assessed whether the absolute plausibility and the relative plausibility of the role-reversed interpretation predicted

accuracy on the comprehension questions. Response times to the questions were also analyzed as a secondary dependent measure. We do not have a priori predictions about the response time data because the current study is primarily concerned with the accuracy of sentence interpretation rather than the time required to construct them. Indeed, a core goal of this research program is to put the focus on the content of people's interpretations rather than on the time required to construct them.

2.1 Participants

Two hundred fifty-nine undergraduate students participated in the experiment for course credit. Data from 19 participants were excluded because they did not meet the a priori criteria for data inclusion (see section 2.4 Analysis), thus the analyses were conducted on a total of 240 participants.

2.2 Materials

The experimental stimuli consisted of 144 sentences that described a transitive event. Half of the items were the same sentences used in Ferreira (2003), and seventy-two additional stimuli were created to increase statistical power for the current study. Each sentence varied in its structure (active or passive), and in the order of the arguments (one or the other). By design, these experimental items can be classified into three different types according to the meaning resulting from the order of arguments. Forty-eight of the items were nonreversible because one argument was animate whereas the other was inanimate (e.g., *the mouse ate the cheese / the cheese at the mouse*), thus one of the argument orders was semantically anomalous. Another forty-eight items were reversible, biased items: one of the two argument orders was less plausible than the other, but not completely anomalous (e.g., *the dog bit the man / the man bit the dog*). The remaining forty-eight

items were symmetrical because the two argument orders were equally plausible (e.g., *the customer thanked the clerk / the clerk thanked the customer*).

To obtain the critical measures for analyzing the effect of the role-reversed interpretation's plausibility, we conducted two plausibility norming studies on the experimental items described below.

2.2.1 Absolute plausibility norming

In this first norming study, we collected plausibility ratings for the experimental items from 60 undergraduates (none of whom participated in the main experiment). To avoid misinterpretations, only the active version was presented, under the assumption that the semantic plausibility of the active and passive versions will be approximately equivalent. Initially we constructed 150 experimental items to be normed, and they were counterbalanced across two lists so that participants saw only one argument order of each item. Participants were asked to rate how semantically plausible each sentence was from 1 (Very Implausible) to 7 (Very Plausible), and they were presented with four practice trials describing the scale to ensure that they understood the plausibility rating task. The mean plausibility ratings obtained for each sentence type (illustrated in Figure 1) were as expected. For nonreversible items, the plausible argument order ($M = 6.38$, $SD = 0.84$) was rated as more plausible than the anomalous order ($M = 1.50$, $SD = 1.04$), and this difference was statistically significant ($\beta = 4.88$, $t = 149.29$, $p < .001$). Likewise for biased items, plausible versions ($M = 6.18$, $SD = 1.01$) were rated more plausible than the implausible versions ($M = 2.90$, $SD = 1.50$; $\beta = 3.28$, $t = 76.23$, $p < .001$). For symmetrical items, the two arbitrary argument orders (order 1: $M = 5.55$, $SD = 1.36$; order 2: $M = 5.51$, $SD = 1.39$) were rated as

similarly plausible, as expected. These absolute plausibility ratings were used as continuous predictors in the analyses of the main experiment.

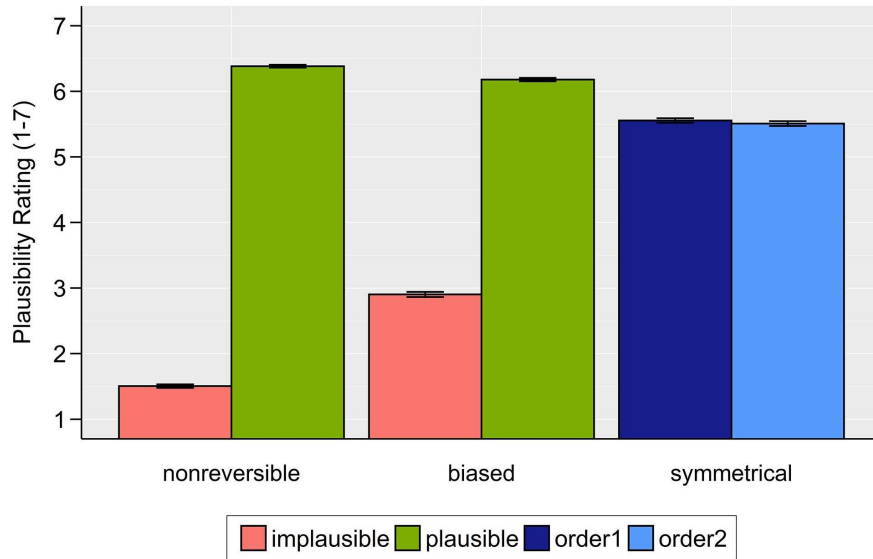


Figure 1. Mean ratings from the absolute plausibility norming study, by sentence type (nonreversible, biased, and symmetrical) and by veridical plausibility (plausible, implausible, order1, and order2). Error bars represent the standard error.

2.2.2 Relative plausibility norming

In addition to the aforementioned norming study that collected the absolute plausibility rating of each sentence (e.g., how plausible is the event *the cat chased the mouse*, relative to all other events in the world?), a second norming study was conducted to collect the relative plausibility of the two argument orders for each sentence type (e.g., how much more plausible is *the cat chased the mouse*, compared to *the mouse chased the cat*?). Another set of 60 undergraduates participated in this study. For each item, participants read the two versions of the experimental item (e.g., *the cat chased*

the mouse; the mouse chased the cat) located on opposite sides of a slider scale with 15 tick marks. The left-side and right-side event were arbitrarily labelled as *A* and *B*, respectively. The scale's left anchor was labelled "A is extremely more likely" and the right anchor "B is extremely more likely", and the middle anchor was labelled "A and B are equally likely". Participants were instructed to read each pair of events, and to drag the slider to rate which event was more likely than the other. The scale ranged from 0 at the midpoint (representing equal plausibility) to 7 demarcations on each side of the slider, to mirror the 7-point Likert Scale from the first plausibility norming study). The possible ratings for each item were integer values from -7 to 7. Moving the slider to the left indicated that the left-side sentence was more likely (and vice versa for the right), and clicking at the center of the scale indicated that the two events are equally likely. The location of each sentence version was randomized so that the plausible and implausible sentences appeared equally often on the left and right sides. Participants were familiarized with the task in four practice trials describing the scale, and then they rated the 150 experimental items presented in random order.

The mean relative plausibility ratings are illustrated in Figure 2. We assessed whether the mean relative plausibility ratings across items corresponded to the criteria defining each sentence type. Based on the mean relative plausibility ratings, we excluded six items from the main experiment because they deviated the most from the following criteria. For the nonreversible items, we excluded two items that had the lowest mean relative plausibility ratings for the implausible items, because those ratings suggested that the items were more reversible than expected. For the symmetrical items, we excluded two items whose mean relative plausibility rating deviated the most from zero, which indicated that the relative plausibility of those event pairs was not equivalent as intended. For the biased items, we excluded two items whose mean relative

plausibility rating was closest to zero, which suggested that those event pairs were not as biased as expected. For the remaining 144 items, the mean relative plausibility ratings were uniform within each sentence type. The mean relative plausibility rating across the nonreversible items indicated that the more plausible versions were considerably more likely than the less plausible versions (using the less plausible items as the baseline; $M = 6.21$, $SD = 2.08$). For the biased items, the mean relative plausibility rating showed that those items were reversible but biased, as intended (compared to the less plausible items as the baseline; $M = 4.98$, $SD = 2.90$). For the symmetrical items, we used the item presented on the right side of the scale as the baseline, since by design the two versions are equally plausible. The mean plausibility ratings ($M = 0.04$, $SD = 2.72$) confirmed that those items were rated as equally plausible. This relative plausibility rating study indicated that these experimental items had the appropriate semantic properties for the experiment, and these relative plausibility ratings were used as predictors in the main experiment.

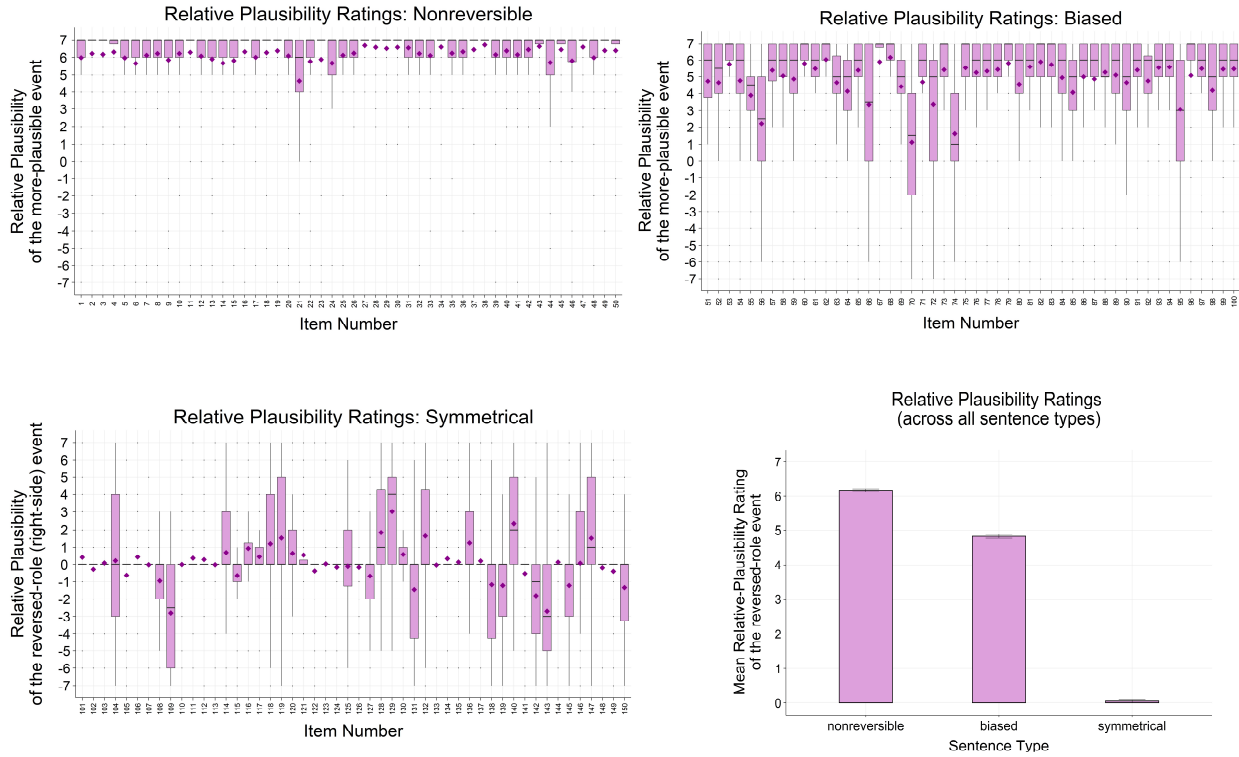


Figure 2. Mean ratings from the relative plausibility norming study, by sentence type (top left: nonreversible; top right: biased; bottom left: symmetrical) and across all sentence types (bottom right). Error bars represent the standard error.

In the main experiment, the 144 experimental items were presented with 100 fillers of various syntactic constructions (for more details, see section 2.3 Procedure). As in Ferreira’s (2003) study, 80 of the filler items were designed so that the sentence can be probed regarding either the location, time, action, or the color of a mentioned object (e.g., *the man snoozed the red alarm clock this morning*). The remaining 20 fillers were constructed as transitive, passive sentences so that they can be probed about the patient (e.g., *the dress was sewn by the man*). This additional filler type was included because unlike in Ferreira (2003), all the experimental items in the current study were only probed regarding the agent but not the patient. The 144 experimental items were

counterbalanced across four lists in a Latin Square design so that each participant read only one version of each experimental item, and so they read an equal number of active and passive sentences as well as an equal number of plausible and implausible sentences. Participants were randomly assigned to one of these lists, and they read and answered comprehension questions on a total of 244 items.

2.3 Procedure

The experiment was designed and presented online using the PC Ibex Farm platform (<https://expt.pcbex.net/>). The experiment began with instructions informing participants that they would be reading individual sentences and answering a question about each one. The six question types were described and illustrated with an example sentence: 1) DO-ER? (corresponding to agent): Who is doing the action in the sentence? For instance, in the sentence *the police arrested the criminal*, the do-er would be the police because they are the entity doing the action. 2) ACTED-ON (corresponding to patient): Who is the action being done to? In the same sentence, the answer is *criminal* because it is the entity being acted-on by the do-er. 3) ACTION?: What is the action in the sentence? For that same sentence, the correct answer would be *arresting* or *arrested*. 4) WHEN?: When did the event occur? In *yesterday the boy flew the kite at the park*, the correct response would be *yesterday*. 5) WHERE?: Where did the event occur? In the same sentence, the correct answer would be *park* or *at the park*. 6) COLOR?: What is the color of the mentioned object? In *the actress drove the red car to the party*, the correct answer would be *red*.

Participants were instructed to read each sentence at a normal pace and to press a spacebar after they were done reading, which would replace the screen with one of the six question probes and a textbox directly below. After typing their answer into the textbox, they pressed Enter to

submit their response and advance to the next trial. In between trials, a buffer screen displayed a message reminding participants to press the spacebar to see the next sentence, and participants were informed that they could take a break during these between-trial periods if desired. To ensure that they understood the probes and were familiarized with the task, participants read six sentences in a practice block with feedback describing the correct response. Afterwards, the 244 items were presented in random order, without feedback on the responses. Each experimental session lasted about 60 minutes.

2.4 Analysis

2.4.1 Data exclusion

Following the a priori criteria we had established for data exclusion, we removed data from six participants whose comprehension accuracy on the filler items was below 85%, using the following procedure. We excluded the 20 fillers asking about the acted-on (patient) because they were potentially more difficult to answer than the other probes on the filler items, and their inclusion would lead to stringent data exclusion. Because the text responses contained numerous suspected misspellings and typographical errors, the following data cleaning procedure was then performed in R (version 3.6.3) to obtain a more precise calculation of accuracy on the remaining 80 fillers. All responses were transformed into lowercase letters, and extra whitespaces and non-letter characters (e.g., . \ ?) were removed. The *hunspell* package was used to perform automated spellchecking and autocorrection. For each misspelling, the first suggested correction from *hunspell* was applied to the responses (e.g., *aborsbed* changed to *absorbed*). Discrepancies between British and American spellings across responses were manually corrected (*grey* changed to *gray*) so that both variants would be matched as correct. Similarly, responses to the action probe were lemmatized

(e.g., *embraced* changed to *embrace*) using the *textstem* package so that different forms of the correct verb would be matched as correct. To calculate the filler accuracy, we used the *agrep* function to apply a fuzzy matching algorithm between the cleaned responses and the correct answers. If the response string (e.g., *in the hallway* or *the hallway*) contained the content word in the correct answer (*hallway*), the response was matched as correct. Mean accuracy on the fillers was high ($M = 96.79\%$). However, six subjects performed below 85%, thus their data was excluded in the main analyses.

To ensure that the data were fully counterbalanced across the four lists, we also removed data from 13 additional subjects (i.e., those who participated after the planned sample size of 60 participants per condition was reached). All subsequent analyses were conducted on the remaining 240 subjects only. One item from the biased sentence type was excluded from the analyses due to coding error.

2.4.2 Data preparation

Because the responses on the experimental items also contained numerous suspected misspellings and typographical errors that most likely reflected errors in typing and spelling rather than a failure to identify the agent of the sentence, we applied a similar data cleaning procedure to obtain a more precise calculation of the mean question answering accuracy on the experimental items. All responses were transformed into lowercase letters, and extra whitespaces between words and non-letter characters were removed. Determiners (e.g., *a*, *an*, *the*) were removed so that only the subsequent content word will be matched against the correct answer (e.g., *man* vs. *the man*). Potentially misspelled words were automatically corrected using the *hunspell* package. We also

manually corrected prevalent misspellings that were not captured by *hunspell* (e.g., *solider* changed to *soldier*, and *principle* changed to *principal*).

In some cases, the algorithm from *hunspell* had corrected the response into a word that was not from the sentence (e.g., *agg* was changed to *gag*, but the correct answer was *egg*). To account for these remaining potential misspellings, we used the fuzzy matching algorithm from the *agrep* function to calculate the mean accuracy. The algorithm is based on the generalized Levenshtein edit distance, which considers the minimal possibly weighted number of insertions, deletions, and substitutions needed to transform the target string (here, the cleaned responses) into the pattern string (here, the correct answers). In setting the algorithm's parameters, we prioritized minimizing false positives (incorrect responses being matched as correct) over reducing these false negatives (potential misspellings not being matched as correct). We set the *max.distance* argument to be 0.05 (default=0.1) to specify a relatively limited number of transformations that are allowed on the pattern. There is no single *max.distance* parameter that will work well for all input strings, but the chosen parameter was optimal for the similarity between arguments unique to our set of items. Using this parameter, suspected typographical errors and misspellings (all within 1 edit distance) that were not correctly changed by *hunspell*'s first suggestion were fuzzy-matched as correct (e.g., *layer-lawyer*, *rabbi-rabbit*), and plural variants were fuzzy-matched as correct (e.g., *troops-troop*). Importantly, these parameters ensured that incorrect responses denoting the other argument in the sentence (e.g., *cow-cud*, *man-woman*, *actor-actress*) were not matched, despite their length and substring similarity. In total, 209 responses were fuzzy-matched as correct (10% of the incorrect responses that were calculated via exact string match).

To calculate the response times to the correct answers, for each trial we subtracted the time when subjects pressed a button to view the question from the time when subjects typed the first letter as their response. This measure of response time accounts for differences in typing speed between subjects because it calculates the response time to type the first letter, rather than the time taken to type the submitted response. Following the data exclusion criteria used in Ferreira's (2003) study, we excluded response times shorter than 300ms and longer than 7000ms. This affected less than 1% of the data. Because preliminary analyses on the raw response times revealed that the assumption of homoscedasticity was not met, the response times were log transformed in the analyses reported below. Response times to the incorrect answers were not analyzed because syntactic or semantic revisions would presumably take place only on the accurate trials.

2.5 Results: a priori analyses

Mean accuracy and response times by conditions for each sentence type are illustrated in Figure 3. Overall, accuracy was remarkably high (<92%) across all conditions. The response times were overall faster than what was reported in Ferreira (2003), which may be partly due to the different methods in calculating them. The response times in the current study were calculated as time taken to type the first letter of the response after the onset of the probe, whereas in Ferreira (2003) it was calculated as the time taken to produce an oral response. Because this difference precludes a direct comparison of the response times data to what was reported in Ferreira (2003), and because they were not our primary dependent measure, we do not provide corresponding figures for them.

In the analyses reported below, each sentence type was analyzed separately (nonreversible, biased, and symmetrical). Accuracy data were analyzed using logistic mixed-effects models, whereas

the response time data were analyzed using linear mixed effects models, both conducted using the *lme4* package (version 1.1-21) in R (version 3.6.3). All models included by-subject and by-item random intercepts and slopes for all the predictor variables and all interactions between them. If a model produced a warning for singular fit, we iteratively removed the random effect that had the smallest variance until the fit was no longer singular. The same approach was followed if a model failed to converge. Model diagnostics were inspected to ensure that the assumptions for linear regression were met.

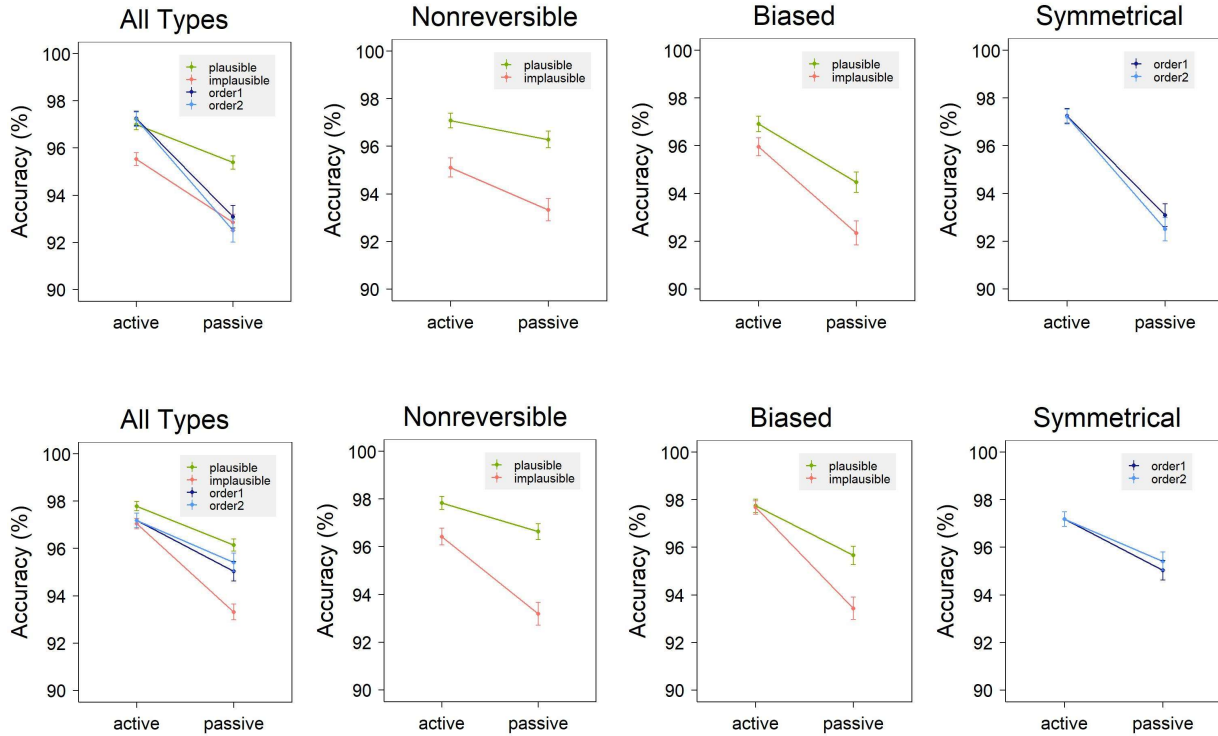


Figure 3. Percentage of correct answers by structure and by veridical plausibility, for the reading comprehension study (Experiment 1; top) and the aural comprehension study (Experiment 2; bottom). Error bars show the standard error.

2.5.1 Effect of structure and veridical plausibility: accuracy

We first examined the hypothesis regarding the effects of structure and veridical plausibility. We entered the interaction between the structure (active or passive) and veridical plausibility (a dichotomous predictor; plausible/order1 or implausible/order2) as fixed effects, and accuracy on the experimental items as the dependent measure. To facilitate model convergence, the predictors were mean-centered and deviation-coded. This procedure was followed for the analyses of the response times as well.

Nonreversible items. For the nonreversible items, there was a significant effect of veridical plausibility, indicating that accuracy was lower for implausible sentences than plausible sentences

($\beta = -0.63, p < .001$). Neither the effect of structure nor the interaction of structure and veridical plausibility were significant (for structure, $\beta = -0.20, p = .167$; for the interaction, $\beta = -0.08, p = .746$).

Biased items. For the biased items, the model revealed significant effects for structure ($\beta = -0.68, p < .001$) and veridical plausibility ($\beta = -0.38, p = .005$), indicating that accuracy was lower for passive than active sentences, and that accuracy was lower for implausible sentences than plausible sentences. The interaction of structure and veridical plausibility was not significant ($\beta = -0.10, p = .620$).

Symmetrical items. For the symmetrical items there was a significant effect of structure, indicating that accuracy was poorer for passive than active sentences ($\beta = -1.08, p < .001$). The effect of veridical plausibility was not significant ($\beta = -0.06, p = .643$), as expected, and neither was the interaction of structure and veridical plausibility ($\beta = -0.12, p = .601$).

2.5.2 Effect of structure and veridical plausibility: response times to correct answers

Nonreversible items. In contrast to the analysis on accuracy, for the nonreversible items there was a significant effect of structure on response times, indicating that response times were longer for passive sentences than active sentences ($\beta = 0.06, p < .001$). The effect of veridical plausibility was not significant ($\beta = -0.01, p = .427$). There was also a significant interaction of structure and veridical plausibility ($\beta = -0.06, p = .002$). Follow-up analyses using Tukey's method for multiple comparisons revealed that response times were longer in the passive plausible condition than the active conditions (p 's $< .001$). In addition, response times were longer for passives than actives when the sentence was plausible ($p < .001$), but not when the sentence is implausible (p 's $> .1$). The remaining contrasts were not significant. Overall, the pattern suggests

that the effect of structure on response times is driven by the longer response times in the passive plausible condition in particular, compared to the active conditions.

Biased items. The response time results for the biased items were similar to the accuracy results. There was a main effect of structure ($\beta = 0.10, p < .001$) and veridical plausibility ($\beta = 0.03, p = .006$), indicating that response times were longer for passive than active sentences, and that responses times were longer for implausible than plausible sentences. The interaction of structure and plausibility was not significant ($\beta = -0.03, p = .150$).

Symmetrical items. Consistent with the accuracy results, for the symmetrical items the only significant effect was that of structure ($\beta = 0.12, p < .001$), indicating that response times were slower for passive than active sentences. The effect of veridical plausibility and the interaction of structure and veridical plausibility were not significant (for veridical plausibility, $\beta = -0.004, p = .711$; for the interaction, $\beta = -0.01, p = .512$).

2.5.3 Effect of structure and the competing interpretation's plausibility: accuracy

Next, we examined our main hypothesis regarding the influence of the competing interpretation's plausibility on comprehension. Mean accuracy by structure and the role-reversed plausibility and relative plausibility variables are illustrated in Figure 4. The following procedure was used for the analyses of the accuracy data, as well as the analyses of response time data which we report for completeness. We entered the interaction between structure and one of the two competing interpretation variables obtained from the norming studies (either role-reversed plausibility or relative plausibility) as fixed effects. These two plausibility predictors were analyzed in separate models because the predictors were highly correlated (except for the symmetrical sentences, which did not have high collinearity between the two predictors). The predictors were

mean-centered, except for relative plausibility because this variable had a meaningful zero value denoting equivalent plausibility ratings between the two sentence versions. Consistent with the analyses reported in Sections 2.5.1 and 2.5.2, structure was deviation coded. The competing interpretation predictors were also scaled using the *scale()* function to prevent model nonconvergence. Although the role-reversed plausibility and relative plausibility predictors were analyzed in separate models, their results will be discussed together because they revealed similar findings.

Nonreversible items. The model with role-reversed plausibility revealed a significant main effect of role-reversed plausibility ($\beta = -0.32, z = p < .001$), indicating that accuracy was lower as the absolute plausibility of the non-veridical, role-reversed meaning increased. Neither the effect of structure nor the interaction of structure and role-reversed plausibility were significant (for structure, $\beta = -0.21, z = p = 0.15$; for the interaction, $\beta = -0.03, z = p = .845$). The model with relative plausibility revealed a significant effect of relative plausibility ($\beta = -0.32, p < .001$), indicating that accuracy was poorer as the relative plausibility of the role-reversed meaning increased. The main effect of structure and the interaction of structure and relative plausibility were not significant (for structure, $\beta = -0.20, z = p = 0.16$; for the interaction, $\beta = -0.04, z = p = .731$).

Biased items. Consistent with the analysis that used veridical plausibility as a predictor, in the model with role-reversed plausibility there was an effect of structure ($\beta = -0.69, p < .001$) and role-reversed plausibility ($\beta = -0.19, p = .006$), indicating that accuracy was poorer for passive than active sentences, and that accuracy was poorer as the absolute plausibility of the role-reversed meaning increased. The interaction of structure and role-reversed plausibility was not significant (β

= 0.02, $p = .821$). Results from the model with relative plausibility were similar: there was an effect of structure ($\beta = -0.68$, $p < .001$) and relative plausibility ($\beta = -0.19$, $p = .005$), indicating that accuracy was lower for passives than actives, and that accuracy was lower as the relative plausibility of the role-reversed meaning increased.

Symmetrical items. The separate analyses of role-reversed plausibility and relative plausibility revealed consistent results: the only significant effect was that of structure (p 's < .001), indicating lower accuracy for passive than active sentences. There was no significant effect of role-reversed plausibility ($\beta = -0.02$, $p = .741$), relative plausibility ($\beta = -0.05$, $p = .508$), nor the interaction of structure and either of these plausibility predictors (p 's > .2).

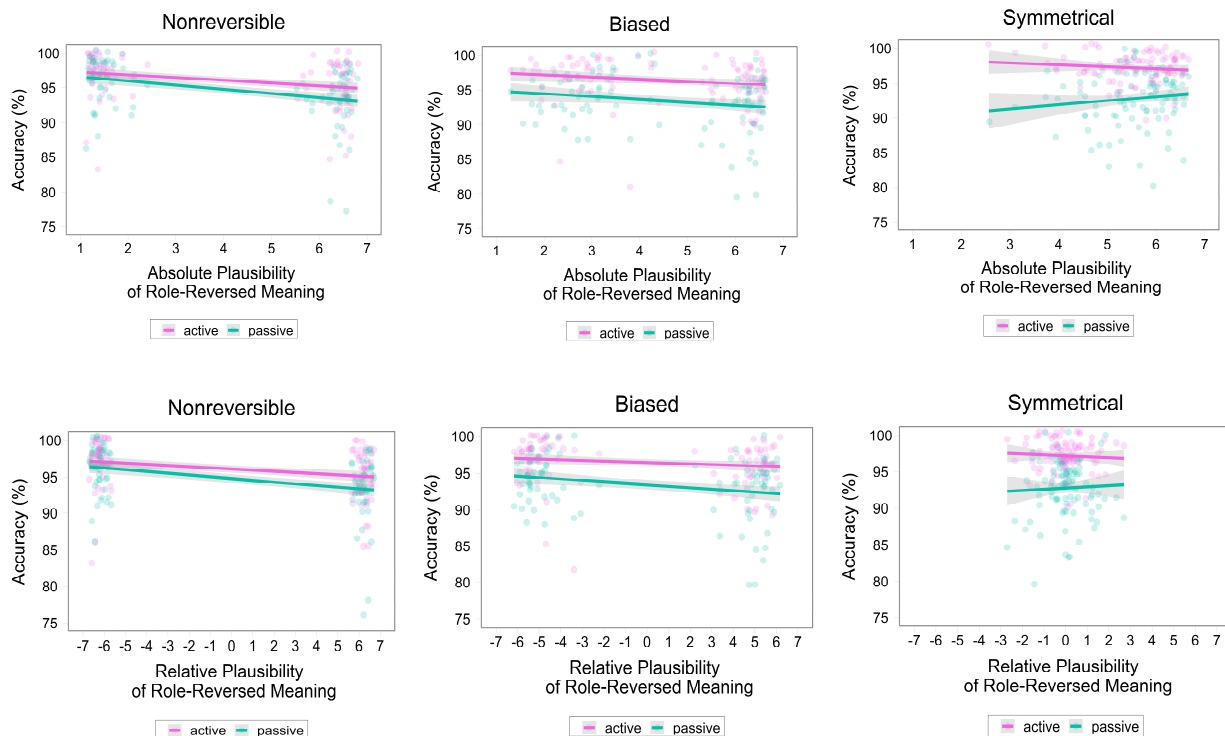


Figure 4. Percentage of correct answers in Experiment 1, by structure and role-reversed plausibility (top), and by structure and relative plausibility (bottom). Error bars show the standard error.

2.5.4 Effect of structure and the competing interpretation's plausibility: response times to correct answers

Nonreversible items. The model that included role-reversed plausibility as a predictor of response times revealed a significant effect of structure ($\beta = 0.06, p < .001$), whereas the effect of role-reversed plausibility was not significant ($p = .454$). There was also a significant interaction of structure and role-reversed plausibility ($\beta = -44.89, p < .001$), indicating that there is a more negative relationship between role-reversed plausibility and response times for passive than for active sentences. The model with relative plausibility revealed a significant effect of structure ($\beta = 0.06, p < .001$) as well as an interaction of structure and relative plausibility ($\beta = -44.06, p < .001$), indicating that response times are slower as the relative plausibility of the role-reversed meaning increases, and that this effect is observed for the active but not the passive sentences.

Biased items. For the biased items, the model with role-reversed plausibility revealed significant main effects for structure ($\beta = 0.10, p < .001$) and role-reversed plausibility ($\beta = 0.02, p = .006$), indicating that response times were slower for passive than active sentences, and that response times were slower as the absolute plausibility of the role-reversed meaning increased. Results from the analyses using relative plausibility were consistent with these findings. There were significant effects of structure ($\beta = 0.10, p < .001$) and role-reversed plausibility ($\beta = 0.02, p = .007$), indicating that response times were longer for passive than active sentences, and that response times were longer as the relative plausibility of the role-reversed meaning increased. The interaction effect was not significant in either model (p 's $> .1$).

Symmetrical items. For the symmetrical items, the model with role-reversed plausibility revealed a significant effect of structure ($\beta = 0.12, p < .001$), indicating that response times were longer for passive than active sentences. The effect of role-reversed plausibility was not significant

($\beta = 0.007, p = .444$), and neither was the interaction between structure and role-reversed plausibility ($\beta = 0.02, p = .089$). For the model with relative plausibility, on the other hand, there was a significant main effect of structure ($\beta = 0.12, p < .001$) as well as relative plausibility ($\beta = 0.01, p = .027$), indicating that response times were longer for passives than actives, and that response times were longer as the relative plausibility of the role-reversed meaning increased. The interaction between structure and relative plausibility was not significant ($\beta = 0.0007, p = .94$).

2.6 Results: exploratory analyses

2.6.1 All sentence types: effect of competing interpretations on accuracy

In addition to analyzing the effect of the competing interpretation for each sentence type separately (nonreversible, biased, and symmetrical items), we conducted an exploratory analysis to investigate whether the role-reversed and relative plausibility predictors accounted for comprehension accuracy when all three sentence types were analyzed in the same model. Mean accuracy for all sentence types by structure and the role-reversed plausibility and relative plausibility variables are illustrated in Figure 5. This aggregated analysis can reveal how the full scale of the competing interpretation's plausibility ratings can account for the rate of misinterpretations, since in this analysis there is data populated throughout the full range of the scale (1 to 7 for absolute plausibility, and -7 to 7 for relative plausibility). As with the a priori analyses, role-reversed plausibility and relative plausibility were tested in separate models. Each model included the interaction between structure and one of the competing interpretations as fixed effects, and the dependent measure was the accuracy data on all the experimental items.

For the model with role-reversed plausibility as a predictor, there were significant main effects of structure ($\beta = -0.62, p < .001$) and role-reversed plausibility ($\beta = -0.23, p < .001$),

indicating that overall accuracy was lower for passive than active sentences, and that accuracy was also lower as the absolute plausibility of the role-reversed meaning increased. The model with relative plausibility revealed the similar findings. There were significant effects of structure ($\beta = -0.62, p < .001$) and relative plausibility ($\beta = -0.20, p < .001$), indicating that accuracy was lower for passive than active sentences, and that accuracy was lower as the relative plausibility of the role-reversed meaning increased. In both models, the interaction effect was not significant (p 's $> .5$).

Overall, this exploratory analysis indicated that when the experimental items were collapsed across sentence type, question answering accuracy was lower for sentences that had a plausible role-reversed interpretation, as well as for sentences whose role-reversed interpretation was highly more plausible than the veridical interpretation.

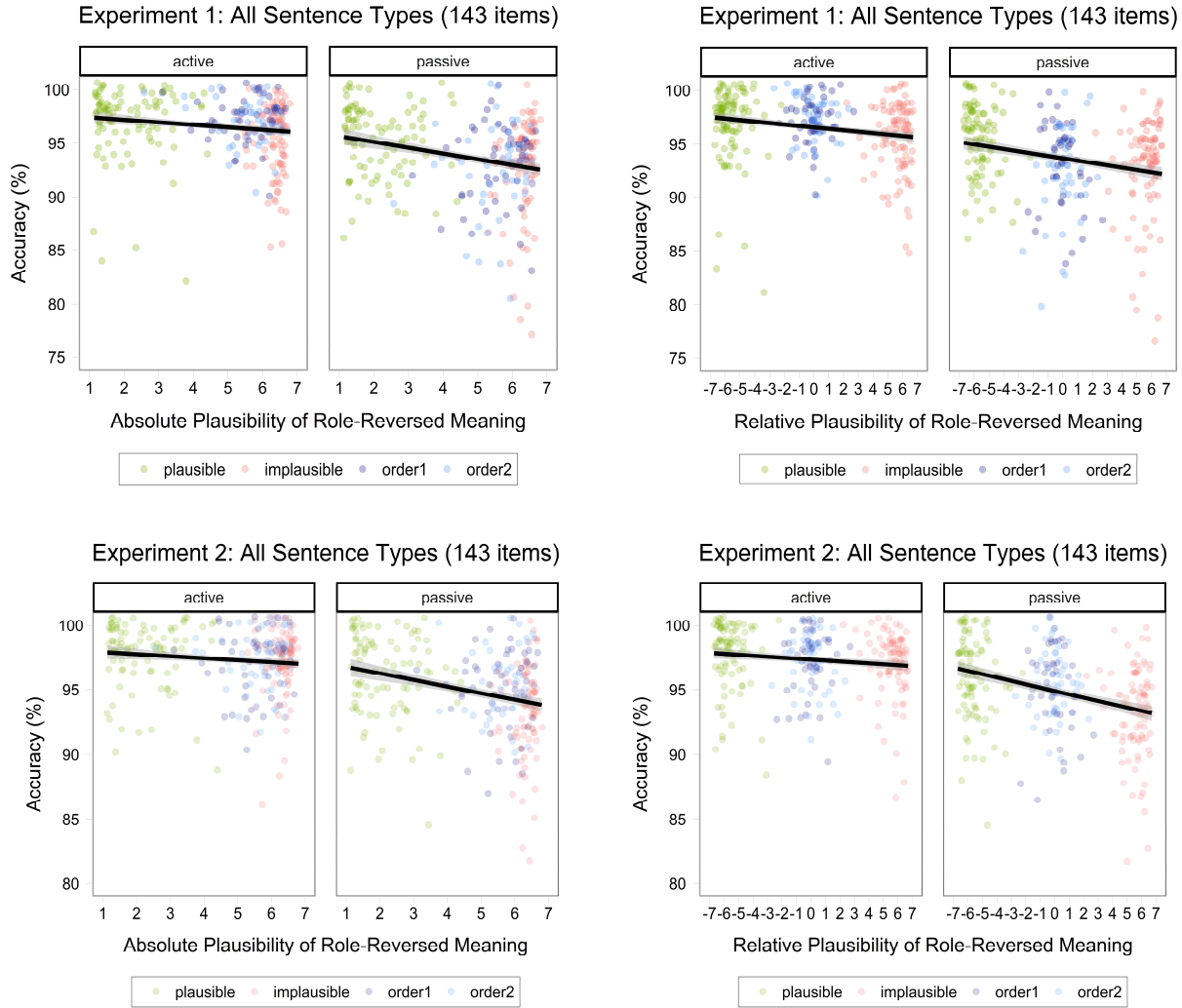


Figure 5. Percentage of correct answers for all sentence types by structure and role-reversed plausibility (left) and relative plausibility (right), in Experiment 1 (top panel) and Experiment 2 (bottom panel). Error bars show the standard error.

2.6.2 Passives only: effect of competing interpretations by sentence type

Since we expected that the role-reversed interpretation would influence the comprehension of passive sentences in particular, in this analysis we examined the effect of competing interpretations on the passive sentences only, and we also tested how this effect might vary

between the three sentence types. Mean accuracy on the passive sentences by structure and the role-reversed plausibility and relative plausibility variables are illustrated in Figure 6. We constructed a linear mixed effects model with the interaction between the sentence type (nonreversible, biased, and symmetrical) and the competing interpretations variable (role-reversed plausibility or relative plausibility) as fixed effects. As with the previous analyses, role-reversed plausibility and relative plausibility were analyzed in separate models.

For the model with role-reversed plausibility there was a main effect of sentence type ($\beta = 0.52, p < .001$), indicating that accuracy was overall higher for nonreversible passives ($M = 94.8\%$) than symmetrical passives ($M = 92.6\%$), but accuracy did not significantly differ between symmetrical passives and biased passives ($M = 93.4\%$) did not significantly differ ($\beta = 0.01, p = .96$). Crucially, there was also a significant interaction between sentence type and role-reversed plausibility, indicating that the effect of role-reversed plausibility on accuracy varied by sentence type ($\beta = -0.43, p = .006$). Follow-up analyses using Tukey's test revealed that the relationship between role-reversed plausibility and accuracy is more negative for the nonreversible passives compared to the symmetrical passives ($\beta = -0.50, p = .002$) and the biased passives (trending effect; $\beta = -0.27, p = .06$). The model with relative plausibility mirrored these findings. There was a significant main effect of sentence type ($\beta = 0.61, p < .001$), indicating that accuracy was overall higher for nonreversible passives than symmetrical passives ($\beta = 0.61, p < .001$). There was also a significant interaction between sentence type and relative plausibility ($\beta = -0.40, p = .04$). Follow-up tests revealed that the relationship between relative plausibility and accuracy is more negative for the nonreversible passives compared to the symmetrical passives ($\beta = -0.54, p < .001$) and the biased passives ($\beta = -0.37, p = .007$).

In sum, this exploratory analysis suggests that the plausibility of role-reversed interpretation incurred the largest cost on question answering accuracy in the nonreversible passive sentences, compared to the biased passive sentences and the symmetrical passive sentences. This was evident for both the role-reversed plausibility and the relative plausibility variables, although the effect was more robust when relative plausibility was used as a predictor.

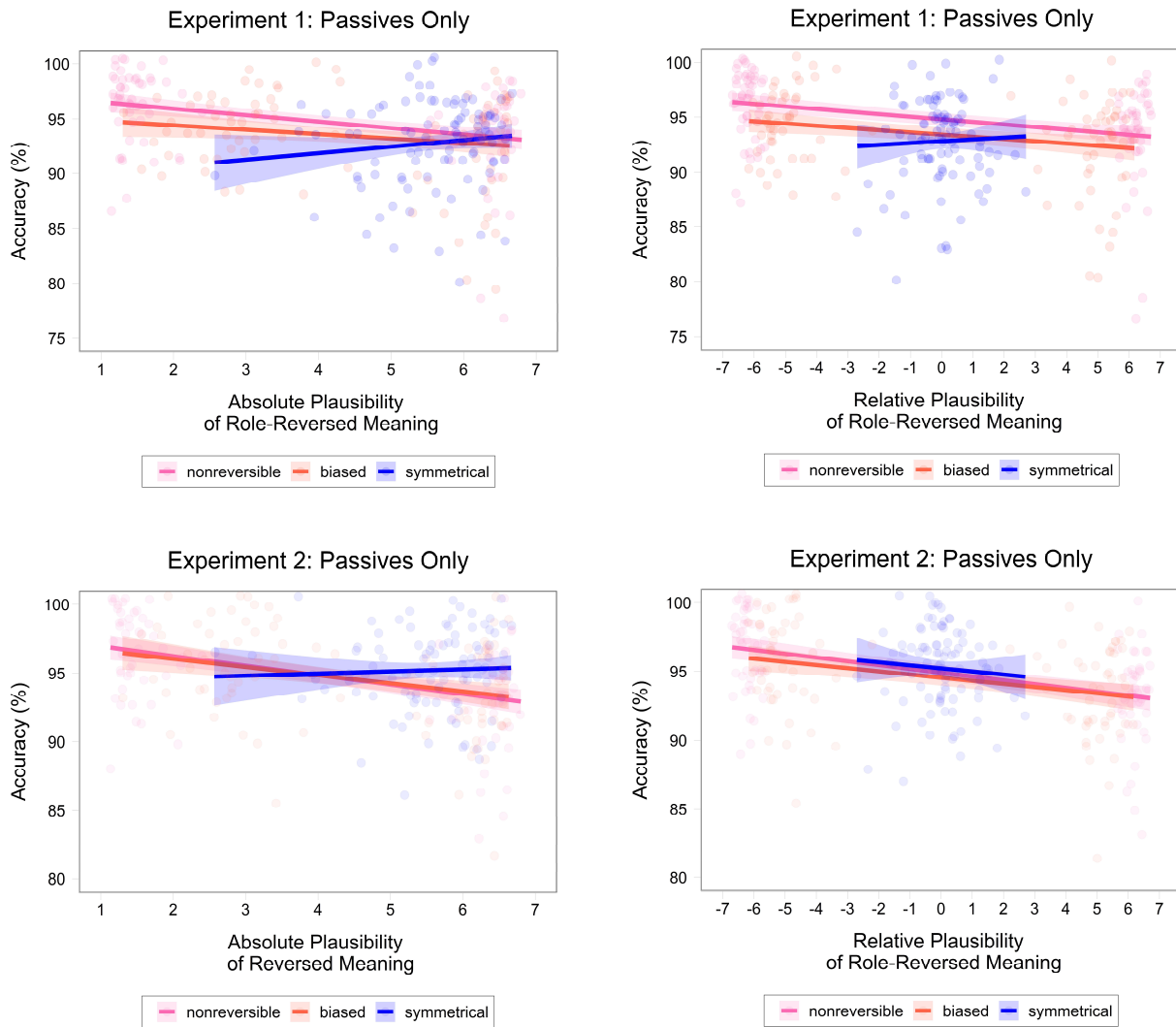


Figure 6. Percentage of correct answers on the passive sentences only, by role-reversed plausibility and sentence type (left) and by relative plausibility and sentence type (right), for Experiment 1 (top panel) and Experiment 2 (bottom panel). Error bars show the standard error.

2.7 Discussion

The main findings from Experiment 1 suggested that the plausibility of the role-reversed interpretation accounts for the comprehension of the nonreversible and biased passive sentences, but not the symmetrical passive sentences. For the nonreversible and biased sentences there was a significant negative relationship between accuracy and the role-reversed interpretation, indicating that the more plausible the role-reversed interpretation, the more errors people made on the agent identification question. However, for the symmetrical sentences, neither measure of the role-reversed interpretation's plausibility accounted for comprehension accuracy, even though the two plausibility norming studies had shown that these symmetrical sentences also had a compelling, alternative interpretation.

Taken together, the results across the three sentence types provide evidence against the view that misinterpretations are driven by a concurrent competition between the veridical interpretation and the role-reversed interpretation. Rather, the process may be more serial, in which the syntactically-mandated parse is restructured only if the veridical interpretation is implausible or anomalous, and this revision occurs only when there is an alternative set of thematic roles that is more plausible than the veridical interpretation. The exploratory analysis further indicated that when all sentence types were included in the model, both measures of the role-reversed interpretation's plausibility predicted comprehension accuracy, consistent with the competing interpretations hypothesis. However, nuanced differences emerged in the exploratory analysis on the passive sentences categorized by sentence type (nonreversible, biased, and symmetrical). Higher plausibility of the role-reversed interpretation predicted lower question answering accuracy for the nonreversible passive sentences, but this effect was weaker for the

biased passive sentences and was completely absent for the symmetrical passive sentences. Visual inspection of the graphs in Figure 4 suggests that there was more error variance in the estimated slope of the role-reversed plausibility and relative plausibility variables for the symmetrical passives and the biased passives, compared to the nonreversible passives, which may explain why the effect was either less robust or absent for those two sentence types. The finding that the competing interpretation's plausibility did not impact accuracy on the symmetrical passive sentences corroborates the results and conclusions from the a priori analyses.

The findings from the analyses using the plausibility of the veridical interpretation as a dichotomous predictor replicated the findings from prior research (Ferreira, 2003; Amichetti, White, & Wingfield, 2016; Dabrowska & Street, 2006). As in these previous studies, people were more likely to misinterpret implausible sentences and passive sentences, compared to plausible sentences and active sentences, respectively. There are some differences in the findings, however. Unlike in Ferreira (2003), we did not find a significant two-way interaction between structure and plausibility. Although the interaction was not significant, the descriptive statistics indicated that accuracy was poorest in the passive implausible condition (see Table 1) compared to the other three conditions, which is consistent with the question accuracy reported in Ferreira's (2003) study. Additionally, accuracy was not affected by passivization for the nonreversible sentences, contrary to our predictions and past research (Ferreira 2003; but see Meng & Bader 2018, 2021). For these nonreversible sentences, the passive structure may not affect accuracy of comprehension because there is only one legitimate assignment of the thematic roles (Slobin, 1966). Participants may also be relying on the animacy cue to respond because the nonreversible items contained one

animate and one inanimate argument (e.g., *The cheese was eaten by the mouse*), whereas this animacy cue is not available in the biased sentences and the symmetrical sentences.

Another notable difference in the results is that the question accuracy in the current study was overall higher than in Ferreira's study, across all the conditions and sentence types. Previous studies that used the same experimental design and paradigm as Ferreira (2003) also found overall higher accuracy rates (Amichetti, White, & Wingfield, 2016; Dabrowska & Street, 2006). The ceiling performance on question accuracy is problematic for testing our main hypothesis regarding the role of competing interpretations because there is little variation in the accuracy between the experimental conditions. Even for the passive implausible condition, performance was at 92% and 93% for the biased and nonreversible sentence types respectively, which is considerably higher than what was reported in prior research (~75%; Ferreira, 2003; Christianson et al., 2010; Bader & Meng, 2018; Meng & Bader, 2021). These differences are likely due to the methodological differences between the two studies, such as the visual presentation of the sentences rather than aural presentation. The unconstrained, untimed reading paradigm used in the current study allows participants to re-read each sentence as much as they like, which may have led to ceiling performance in accuracy. Thus, to address this limitation of this design, in Experiment 2 we used an aural listening task instead. A listening comprehension paradigm presents the sentences as an ephemeral signal and circumvents the potential issue associated with re-reading.

3. Experiment 2: aural study

In Experiment 2, we assessed how the plausibility of the role-reversed meaning influences the comprehension of passive sentences in an aural comprehension study. This experiment used the same experimental design, stimuli, procedure, and analyses as Experiment 1, except the

sentences were presented aurally rather than in written form. The goal of Experiment 2 is to replicate the findings from Experiment 1 by presenting the stimuli in a different modality. Participants listened to sentences that varied in structure (active, passive) and veridical plausibility (plausible/order1, implausible/order2). After each sentence, participants answered a comprehension question asking who was the agent of the action. Like in Experiment 1, we assessed whether the absolute plausibility and the relative plausibility of the role-reversed interpretation accounted for comprehension accuracy. Response times to the questions were collected as a secondary dependent variable.

3.1 Participants

Two hundred sixty-six undergraduate students from the same subject pool as Experiment 1 participated in this study. These subjects did not participate in Experiment 1 nor in any of the plausibility norming studies reported under section 2.2 Materials. Data from 26 participants were excluded (see section 3.4 Analysis), leaving data from 240 participants that were used in the analyses.

3.2 Materials

The experimental stimuli and filler sentences were the same sentences used in Experiment 1, except each sentence was recorded as an mp3 file using Google Cloud's online text-to-speech interface (<https://cloud.google.com/text-to-speech/>). The sentences were recorded using the Voice name "en-US-Wavenet-C" with the default settings for speed, pitch, and voice type. Silent periods at the beginning and end of each audio track were trimmed. We chose to record the sentences using a speech synthesizer rather than a human speaker in order to control for systematic variations in the stimuli's acoustic properties (e.g., amplitude and pitch) across conditions. The

speech synthesizer yielded fairly naturalistic speech recordings and circumvented the need to cross-splice the NPs, which would have resulted in less naturalistic stimuli.

3.3 Procedure

The experimental procedure was the same as Experiment 1, except for the following changes. Participants were instructed to listen to each sentence and type their response to the comprehension question that appeared after each sentence's playback. At the beginning of each trial, participants saw the text "Listen to this sentence..." presented on screen for 800ms, followed by playback of the audio stimulus. Immediately after the audio playback, participants saw the comprehension probe on screen and typed their response into the textbox, as in Experiment 1. After participants submitted their response, on the next screen they saw the prompt "Audio problem?" and were instructed to press the "Y" key if they experienced issues hearing the audio (e.g., could not hear audio at all or audio sounds choppy, whether due to unstable internet connection or other unforeseen technical difficulties). Participants pressed the spacebar to move onto the next trial if they did not experience any issues, otherwise the trial would automatically end if no key was pressed within 8 seconds. This audio probe allowed us to exclude the rare trials on which participants experienced difficulty hearing the stimuli. Aside from these differences, the experimental procedure was identical to that of Experiment 1.

3.4. Analysis

3.4.1 Data exclusion

Following the same data exclusion criteria and procedure as Experiment 1, we removed data from 11 participants. To ensure that the four lists were fully counterbalanced, data from 15 additional subjects were excluded. All analyses were conducted on the remaining 240 subjects.

Trials on which participants reported an audio problem were excluded from the analyses.

Additionally, one item from the nonreversible sentence type was excluded from the analyses due to audio recording error.

3.4.2 Data preparation

Responses on comprehension questions were cleaned using the same procedure as Experiment 1. Because Experiment 2 presented the stimuli aurally rather than in written form, there were more spelling errors not captured by *hunspell* that had to be manually corrected. On these few trials it was fairly obvious from the response that the participant had correctly identified the agent they heard, but that they either made a typological error (e.g., “bridesmade” for “bridesmaid”, or variant spellings for “chauffeur”) or they misheard the noun due to phonological similarity (e.g., mishearing "writer" or "rioter" for the spoken word "rider", in the item set *The horse threw the rider*). Since we do not want these perceptual and typological errors to systematically inflate the error rate on these particular items, we manually corrected these variant responses to the correct answer before applying the fuzzy matching algorithm on the cleaned responses. There were a total of 315 responses that were fuzzy-matched (18% of the incorrect responses calculated via exact string match).

Due to minor differences in the experimental setup of Experiment 1 and 2, the response times for this experiment had to be calculated slightly differently. Unlike in Experiment 1 where participants pressed a button after they finished reading to view the comprehension prompt, in Experiment 2 participants heard automatic playback of aural stimuli and then immediately saw the comprehension prompt. Thus, to calculate the response times to the correct answers, for each trial we took the timestamp when participants typed the first letter of their response and subtracted

from it the playback time for each sentence, and we also subtracted the 800ms interval when they saw the text "Listen to this sentence..." presented on the screen before each stimulus playback. This calculation gives a rough estimate of the time it took participants to type the first letter on each trial. We note that response times were overall longer in Experiment 2 than Experiment 1, which may be partly due to the different methods of calculating the response times, or it may also be the case that participants actually took longer to respond in this experiment than the first one. As in Experiment 1, we excluded response times shorter than 300ms and longer than 7000ms (less than 2.5% of the data), and the response times were log transformed for the analyses.

3.5 Results: *a priori* analyses

Mean accuracy and response times by conditions for each sentence type are reported in Table 1 and illustrated in Figure 3. All analysis procedures were identical to that of Experiment 1.

3.5.1 Effect of structure and veridical plausibility: accuracy

Nonreversible items. For nonreversible items, there were significant main effects of structure ($\beta = -0.63, p < .001$) and veridical plausibility ($\beta = -0.81, p < .001$), indicating that accuracy was lower for passive than active sentences, and that accuracy was lower for implausible sentences than plausible sentences. The interaction of structure and veridical plausibility was not significant ($\beta = -0.31, p = .19$).

Biased items. For the biased items, there were significant main effects of structure ($\beta = -1.00, p < .001$) and veridical plausibility ($\beta = -0.30, p = .03$), indicating that accuracy was poorer for passive than active sentences, and that accuracy was poorer for implausible sentences than plausible sentences. The interaction between structure and veridical plausibility was not significant ($\beta = -0.47, p = .09$).

Symmetrical items. For the symmetrical items there was a significant effect of structure, indicating that accuracy was lower for passive than active sentences ($\beta = -0.60, p < .001$). The main effect of veridical plausibility was not significant ($\beta = 0.07, p = .579$), as expected, and neither was the interaction of structure and veridical plausibility ($\beta = 0.04, p = .87$).

3.5.2 Effect of structure and veridical plausibility: response times to correct answers

Nonreversible items. For the nonreversible items, the response time results were consistent with the accuracy results. There were main effects of structure ($\beta = 0.10, p < .001$) and veridical plausibility ($\beta = 0.10, p < .001$), indicating that response times were slower for passive sentences than active sentences, and that response times were slower for implausible sentences than plausible sentences. The interaction of structure and veridical plausibility was not significant ($\beta = -0.007, p = .739$).

Biased items. For the biased items, there was a significant effect of structure ($\beta = 0.12, p < .001$) and veridical plausibility ($\beta = 0.04, p < .001$), indicating that response times were slower for passive than active sentences, and that response times were slower for implausible than plausible sentences. The interaction between structure and veridical plausibility was marginally significant ($\beta = -0.04, p = .04$). Follow-up analyses using Tukey's method for multiple comparisons revealed that response times did not differ between the passive plausible and passive implausible conditions ($p = .52$), whereas all the remaining pairwise comparisons between conditions were significant (p 's $< .001$). This suggests that the effect of plausibility on response times (i.e., slower response times to implausible sentences than plausible sentences) is evident in the active sentences but not the passive sentences.

Symmetrical items. Consistent with the accuracy results, for symmetrical items the only significant effect was that of structure ($\beta = 0.15, p < .001$), indicating that response times were slower for passive sentences than active sentences. The effect of structure and the interaction of structure and veridical plausibility were not significant (for veridical plausibility, $\beta = 0.003, p = .81$; for the interaction, $\beta = 0.004, p = .82$).

3.5.3 Effect of structure and the competing interpretation's plausibility: accuracy

Next, we examined our main hypothesis regarding the influence of the competing interpretation's plausibility on comprehension accuracy. Mean accuracy by structure and the role-reversed plausibility and relative plausibility variables are illustrated in Figure 7. The model specifications were identical to Experiment 1. Role-reversed plausibility and relative plausibility were analyzed in separate models that included their interaction with structure (Structure x role-reversed plausibility, Structure x Relative plausibility).

Nonreversible items. For the nonreversible items, the results were consistent with the analyses that used veridical plausibility as a predictor. In the model with role-reversed plausibility, there was main effect of structure ($\beta = -0.62, z = p < .001$) and role-reversed plausibility ($\beta = -0.40, z = p < .001$), indicating that accuracy was lower for passive sentences than active sentences, and that accuracy was lower as the plausibility of the role-reversed meaning increased. Similarly, in the model with relative plausibility there was a main effect of structure ($\beta = -0.63, p < .001$) and relative plausibility ($\beta = -0.40, p < .001$), indicating that accuracy was poorer as the relative plausibility of the role-reversed meaning increased. In both models, the interaction was not significant (p 's $> .1$).

Biased items. For the biased items, the model with role-reversed plausibility revealed a significant main effect of structure ($\beta = -0.99, p < .001$) and role-reversed plausibility ($\beta = -0.16, p$

= .008), indicating that accuracy was lower for passive than active sentences and that accuracy was lower as the plausibility of the role-reversed meaning increased, consistent with the analyses using veridical plausibility. The interaction of structure and role-reversed plausibility was not significant ($\beta = -0.25, p = .09$). Results were similar for the model with relative plausibility. There was a significant effect of structure ($\beta = -0.99, p < .001$) and relative plausibility ($\beta = -0.16, p = .01$), indicating that accuracy was lower for passives than actives, and that accuracy was lower as the relative plausibility of the role-reversed meaning increased. The interaction of structure and relative plausibility was not significant ($\beta = -0.23, p = .1$).

Symmetrical items. For the symmetrical items, both models revealed that the only significant effect was that of structure (p 's $< .001$), indicating that accuracy was lower for passive than active sentences. There was no effect of role-reversed plausibility nor relative plausibility on accuracy (p 's $> .8$), and the interaction was not significant in either model (p 's $> .5$).

Table 1

Mean accuracy and response times on the comprehension questions by condition for Experiment 1 and 2

	Experiment 1 (reading)		Experiment 2 (aural)	
	Accuracy (SD)	RT of first letter typed (SD)	Accuracy (SD)	RT of first letter typed (SD)
Nonreversible sentences				
Actives				
Plausible	97 (17)	1405 (631)	98 (15)	1711 (842)
Implausible	95 (22)	1470 (775)	96 (19)	1965 (1118)
Passives				
Plausible	96 (19)	1580 (798)	97 (18)	1956 (1085)
Implausible	93 (25)	1543 (866)	93 (25)	2155 (1191)
Biased sentences				
Actives				
Plausible	97 (17)	1438 (673)	97 (15)	1777 (896)
Implausible	96 (20)	1537 (795)	97 (15)	1917 (1026)
Passives				
Plausible	95 (23)	1673 (909)	96 (20)	2105 (1152)
Implausible	92 (27)	1707 (955)	93 (25)	2164 (1193)
Symmetrical sentences				
Actives				
Order1	97 (16)	1519 (741)	97 (17)	1891 (999)
Order2	97 (16)	1524 (735)	97 (17)	1893 (1006)
Passives				
Order1	93 (25)	1778 (957)	95 (22)	2218 (1152)
Order2	93 (26)	1762 (961)	95 (21)	2225 (1176)

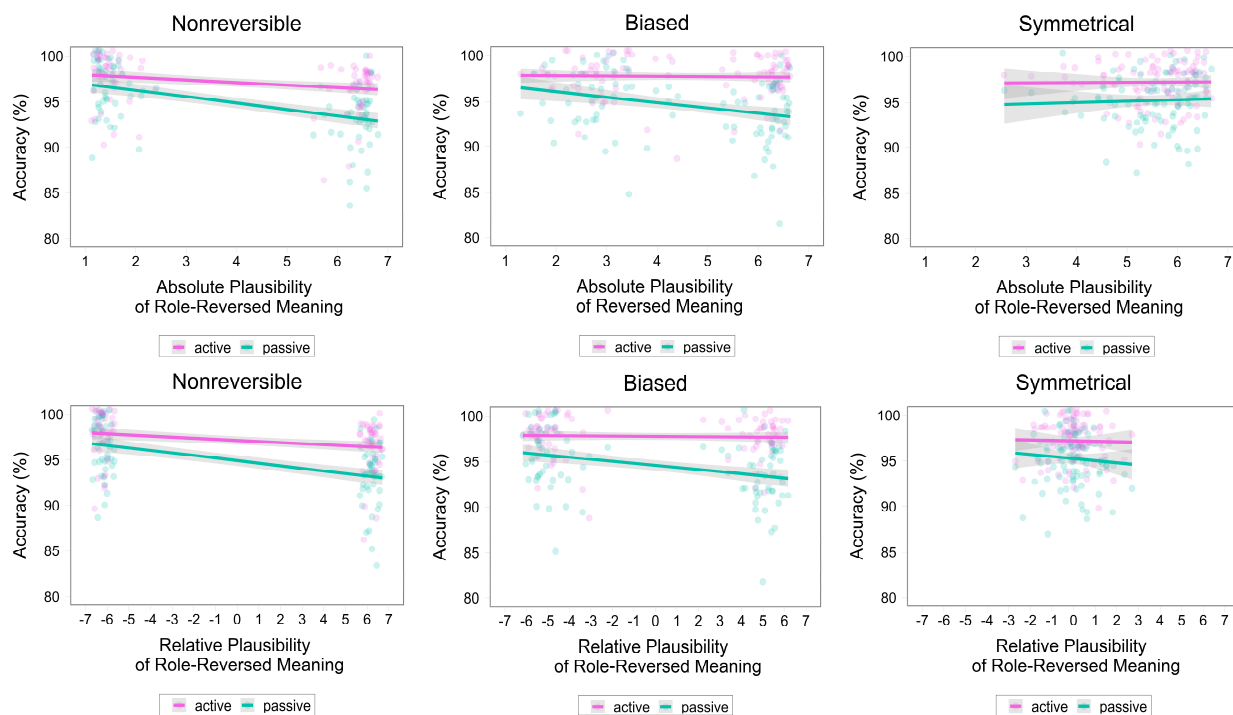


Figure 7. Percentage of correct answers in Experiment 2, by structure and role-reversed plausibility (top), and by structure and relative plausibility (bottom). Error bars show the standard error.

3.5.4 Effect of structure and the competing interpretation's plausibility: response times to correct answers

Nonreversible items. In the model with role-reversed plausibility, there was a main effect of structure ($\beta = 0.10, p < .001$) and role-reversed plausibility ($\beta = 0.05, p < .001$), indicating that response times were slower for passive than active sentences, and that response times were slower as the absolute plausibility of the role-reversed meaning increased. The model with relative plausibility as a predictor revealed similar findings. There was a main effect of structure ($\beta = 0.10, p < .001$) and relative plausibility ($\beta = 0.05, p < .001$), indicating that response times were slower for

passives than actives, and that response times were slower as the relative plausibility of the role-reversed meaning increased. The interaction was not significant in either of these models (p 's > .7).

Biased items. For the biased items, the model with role-reversed plausibility revealed significant main effects of structure ($\beta = 0.12, p < .001$) and relative plausibility ($\beta = 0.02, p < .001$), indicating that response times were slower for passive than active sentences, and response times were slower as the plausibility of the role-reversed meaning increased. There was a trend suggesting an interaction between structure and role-reversed plausibility, but the effect did not reach significance ($\beta = -0.02, p = .07$). Results from the analyses with relative plausibility were consistent with these findings. There was a main effect of structure and relative plausibility, indicating that response times were slower for passive than active sentences ($\beta = 0.12, p < .001$), and response times were slower as the relative plausibility of the role-reversed meaning increased ($\beta = 0.02, p < .001$). The interaction of structure and relative plausibility was marginally significant ($\beta = -0.02, p = .05$).

Symmetrical items. Consistent with the analyses on the symmetrical items using veridical plausibility, the model with role-reversed plausibility and the model with relative plausibility both revealed that the only significant effect was that of structure ($\beta = 0.15, p < .001$), indicating that response times were longer for passive than active sentences. The effects of role-reversed plausibility and relative plausibility were not significant (p 's > .6), and neither were the interactions (p 's > .8).

3.6 Results: exploratory analyses

3.6.1 All sentence types: effect of competing interpretations on accuracy

As in Experiment 1, we conducted an exploratory analysis to test the effect of the competing interpretation's plausibility on comprehension accuracy across all the experimental items, collapsed across sentence types. Mean accuracy for all sentence types by structure and the role-reversed plausibility and relative plausibility variables are illustrated in Figure 5. We ran two separate models for role-reversed plausibility and relative plausibility and included their respective interaction with structure as fixed effects. The model with role-reversed plausibility revealed significant main effects of structure ($\beta = -0.72, p < .001$) and role-reversed plausibility ($\beta = -0.25, p < .001$), indicating that overall accuracy was lower for passive than active sentences, and that overall accuracy was lower as the absolute plausibility of the role-reversed meaning increased. There was a trend suggesting an interaction of structure and role-reversed plausibility ($\beta = -0.12, p = .07$), though this effect did not reach significance. Visual inspection of the graph suggests that for passive sentences, accuracy was lower as the relative plausibility of the role-reversed meaning increased, but this effect may be absent in active sentences. The model with relative plausibility revealed the similar findings. There was a main effect of structure ($\beta = -0.71, p < .001$) and relative plausibility ($\beta = -0.23, p < .001$), indicating that accuracy was lower for passive than active sentences, and accuracy was lower as the relative plausibility of the role-reversed meaning increased. There was also a significant interaction of structure and relative plausibility ($\beta = -0.15, p = .03$). Visual inspection of the graph clearly suggests that for the passive sentences, accuracy was lower as the relative plausibility of the role-reversed meaning increased, and that this effect was absent or weaker for the active sentences.

Overall, consistent with Experiment 1, this exploratory analysis indicated that when all three sentence types were analyzed in the same model, question answering accuracy was lower for sentences that had a plausible role-reversed interpretation, as well as for sentences whose role-reversed interpretation was highly more plausible than the veridical interpretation. Unlike Experiment 1, there was also an interaction effect suggesting that the negative relationship between accuracy and relative plausibility was evident in the passive sentences but not for the active sentences, consistent with our predictions. There was also a trend suggesting an interaction of accuracy and role-reversed plausibility in the same direction as well, although this effect was not statistically significant.

3.6.2 Passives only: effect of competing interpretations by sentence type

Mean accuracy on the passive sentences by structure and the role-reversed plausibility and relative plausibility variables are illustrated in Figure 6. Like in Experiment 1, we also conducted an exploratory analysis to assess the influence of the competing interpretation's plausibility on the passive sentences only, and how this effect might differ by sentence type. In one model we included the interaction of sentence type and role-reversed plausibility as fixed effects, and in a separate model we included the interaction of sentence type and relative plausibility as fixed effects. The model with role-reversed plausibility revealed a main effect of sentence type ($\beta = 0.56$, $p < .001$), indicating that accuracy was overall higher for nonreversible passives ($M = 95.3\%$) than symmetrical passives ($M = 94.7\%$), whereas accuracy for symmetrical passives and biased passives ($M = 94.9\%$) did not significantly differ ($\beta = 0.004$, $p = .2$). Crucially, there was an interaction of role-reversed plausibility and sentence type ($\beta = -0.48$, $p = .002$). Follow-up analyses using Tukey's test revealed that the relationship between role-reversed plausibility and accuracy is more negative

for the nonreversible passives compared to the symmetrical passives ($\beta = -0.57, p < .001$) and the biased passives (marginal effect; $\beta = -0.28, p = .06$). The model with relative plausibility revealed consistent findings. There was a significant main effect of sentence type ($\beta = 0.64, p < .001$), indicating that accuracy was overall higher for nonreversible passives than symmetrical passives. The accuracy for nonreversible passives and biased passives did not significantly differ ($\beta = -0.12, p = .38$). In addition, there was a significant interaction of sentence type and relative plausibility ($\beta = -0.56, p = .004$). Follow-up tests revealed that the relationship between relative plausibility and accuracy is more negative for the nonreversible passives compared to the symmetrical passives ($\beta = -0.58, p < .001$) and the biased passives ($\beta = -0.38, p = .006$).

Overall, this exploratory analysis suggests that the plausibility of the role-reversed interpretation had the greatest impact on question answering accuracy in the nonreversible passive sentences compared to the symmetrical passive sentences, in line with Experiment 1. There was a trending but non-significant effect suggesting that the slope was more negative for the nonreversible passives than the biased passives, which is also consistent with the pattern of results from Experiment 1.

3.7 Comparing Experiments 1 & 2

The goal of Experiment 2 was to test whether the competing interpretations hypothesis accounted for the misinterpretation of passive sentences using a listening comprehension paradigm that addresses some of the limitations of the reading comprehension paradigm used in Experiment 1. If the difficulty in thematic role identification arises due to competition between the veridical interpretation and an alternative, role-reversed interpretation, then the plausibility of this competing interpretation should impact question answering accuracy. As in Experiment 1,

different predictions were derived for the symmetrical sentences in particular, based on the nature of the competition architecture assumed under the hypothesis. Central to our main hypothesis, the results from the a priori analyses replicated the main findings from Experiment 1 in showing that the plausibility of the role-reversed meaning impacted comprehension accuracy on the nonreversible and biased sentences, but not the symmetrical sentences. This provides further evidence against the original formulation of the competing interpretations hypothesis as a potential explanation for the misinterpretation of passive sentences.

Other aspects of the findings were also consistent with Experiment 1. Question answering accuracy was high across all the conditions (see Table 1) in this experiment as well as in Experiment 1. This finding rules out the explanation that the exceptionally high accuracy on the comprehension questions was due to the modality in which the sentences were presented. A possible explanation for the ceiling performance is that the thematic role identification task is overall easier when comprehenders are asked to identify the agent on the majority of the trials as in the two experiments reported here, compared to the task used by Ferreira (2003) in which a higher proportion of the trials asked about the patient of the action. Nevertheless, even if question answering accuracy was exceptionally high, the a priori analyses revealed consistent results across both experiments, and therefore the ceiling performance does not undermine our main conclusions of the present study. The findings from the exploratory analysis on the passive sentences also corroborated the main conclusions, showing that neither role-reversed plausibility nor relative plausibility predicted accuracy when the sentence was symmetrically plausible.

Some differences in the results were observed between the two experiments, however. In this experiment, accuracy on the nonreversible sentences was significantly lower for passive

sentences than active sentences, whereas this effect was clearly absent in Experiment 1. A possible explanation for this difference is that full passives (...*was verbed by*...) are more commonly encountered in written form than in aural form (Roland, Dick, & Elman, 2007), which may explain why we observed a cost for the passive structure when a listening comprehension task was employed (Experiment 2) but not in the reading comprehension task (Experiment 1). In this experiment there was also a significant interaction of structure and the role-reversed plausibility and relative plausibility variables in the exploratory analysis that included all sentence types, an effect that was not observed in Experiment 1. The interaction indicated that the competing interpretation's plausibility had a greater impact on accuracy for the passives than the active sentences, consistent with our prediction. However, because this effect was observed only in Experiment 2 but not in Experiment 1, it is difficult to infer too much from this finding, and further work is required to verify the robustness of this effect. Another notable difference is that response times were overall slower in this experiment than in Experiment 1. This is likely due to the methodological differences in calculating the response times, as described under the Analysis section. There were also several differences in the results from the analyses of response times across the two experiments. For instance, the analyses in Experiment 1 revealed that response times on the symmetrical sentences were significantly slower as the relative plausibility of the alternative interpretation increased, but this effect was absent in Experiment 2, suggesting that it may be a spurious result. Nevertheless, a consistent effect across both experiments is that the response times were reliably slower for passive sentences than active sentences for all three sentence types, indicating a clear response time cost for the passive structure. Because the remaining results from

the analyses of response times were generally not consistent across the two experiments, we do not discuss them further.

4. General Discussion

A goal in language comprehension research is to characterize the mechanisms that underlie thematic role assignment in noncanonical structures such as passive sentences. Prior studies investigating the role of event knowledge in the comprehension of passives have established that there is a greater tendency to misinterpret passive sentences when the veridical interpretation is implausible (Ferreira, 2003; Bader & Meng, 2018; Christianson et al., 2010; Gibson, Bergen, & Piantadosi, 2013). However, it is unclear whether such misinterpretations are driven by competing interpretations in which the parser also considers a non-veridical, role-reversed meaning in assigning thematic roles. To address this theoretical gap, we used continuous measures of event plausibility to assess whether the misinterpretation of passive sentences arises due to concurrent competition between the veridical interpretation and an alternative, compelling. Using the same experimental design and stimuli as Ferreira (2003), we conducted two comprehension studies that employed sentences varying in structure (active or passive) and veridical plausibility (plausible or implausible). If the parser does consider the plausibility of the role-reversed meaning as a source of information in assigning thematic roles to passive sentences, then misinterpretations should be more likely as the plausibility of the role-reversed meaning increases. We assessed this competing interpretations hypothesis in a reading comprehension study (Experiment 1) as well as a listening comprehension study (Experiment 2). Results using the conventional, dichotomous measure of veridical plausibility showed that in both experiments, comprehenders systematically made more

errors on implausible sentences than plausible sentences, and that error rates were also higher for passive sentences than active sentences, replicating prior research (Ferreira, 2003; Amichetti, White, & Wingfield, 2016; Christianson, Luke, & Ferreira, 2010; Meng & Bader, 2018, 2021). More central to the main research question in this study, both experiments also revealed that people made more question answering errors on sentences that had a plausible, role-reversed interpretation, consistent with the competing interpretations hypothesis. Crucially, this effect of role-reversed plausibility on accuracy was observed only for the nonreversible and biased sentences, but not for the symmetrical sentences. In the remaining sections we discuss the implications of our findings on the assignment of thematic roles in passive sentences, focusing on the novel insights provided by the results of the role-reversed plausibility and relative plausibility variables.

Under the competing interpretations hypothesis, we predicted that comprehension accuracy will depend on the absolute plausibility and/or the relative plausibility of the role-reversed meaning. Results for the nonreversible and biased sentences appear consistent with this prediction. Both the role-reversed plausibility and the relative plausibility predictors accounted for comprehension accuracy as well as response times on these two sentence types that contain a syntax-semantics conflict. These effects were also evident in the exploratory analyses that included all three sentence types, indicating that there is a greater tendency to misinterpret sentences especially when there is an alternative interpretation that is more plausible. However, the findings from the symmetrical sentences have to be considered as well. Contrasting predictions were made for the symmetrical sentences in particular, based on the assumptions of parallel processing and serial processing. Unlike the nonreversible and biased items, we found that for the symmetrical items the plausibility of the role-reversed meaning did not affect comprehension accuracy,

suggesting that there is no competition from the alternative, equally plausible meaning. This conclusion is further strengthened by the exploratory analyses on the passive sentences only, which indicated that the competing interpretations effect is absent or reversed in symmetrically plausible passives. The fact that there was no effect of competing interpretations for the symmetrical sentences provides evidence against the competing interpretations hypothesis that assumed automatic activation of the alternative interpretation. If multiple meanings were activated, we expected some competition even when the veridical interpretation and the non-veridical interpretation are equally plausible. Taken together, the findings across all three sentence types do not support the views of parsing that assume automatic activation of more than one interpretation at a time or that there is parallel activation and interaction between all sources of information to determine which interpretation is ultimately selected (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Stevenson, 1994; McRae, Spivey & Tanenhaus, 1998; Levy, 2008; Gibson, Bergen, & Piantadosi, 2013; Hale, 2003).

One possible explanation that can account for data across all three sentence types is that the parser may engage in a serial process of thematic role revision, rather than a concurrent activation of more than one interpretation. Instead of the competing interpretations hypothesis, our findings are more consistent with models of parsing which assume that only one interpretation is adopted at a time, such as revision-as-a-last-resort (Fodor & Inoue, 2000) and the unrestricted race model (Van Gompel, Pickering, & Traxler, 2000, 2001). On the one hand, when the syntactically-mandated parse yields an implausible interpretation, the parser may consider another set of thematic roles and select this non-veridical interpretation. The plausibility of competing interpretation will be important in this case, which can sometimes lead to a wrong decision in the

agent identification task, as observed for the implausible sentences for nonreversible and biased types. On the other hand, when the veridical interpretation of the sentence was plausible, it may not be necessary to consider an alternative set of thematic roles, and therefore there is no revision or competition of the thematic roles. The availability of a plausible, role-reversed interpretation is not sufficient to trigger reanalysis, as indicated by the results from the symmetrical sentences. For thematic revision to take place, the alternative interpretation has to be more plausible than the veridical interpretation. This explanation of our data is consistent with Bader and Meng's (2018; also see Meng & Bader, 2021) studies on the comprehension of passives which found that relative plausibility (quantified as the difference in absolute plausibility between two interpretations) was a stronger predictor of misinterpretations than absolute plausibility. This was also supported by the findings from Amichetti and colleagues (2016) that the lower the plausibility of the veridical interpretation of passive sentences, the more likely they are to be misinterpreted by older adults, suggesting that the plausibility strategy may override full syntactic analysis only when there is a clear syntax-semantics conflict.

The thematic role revision account also garners support from electrophysiological and neuroimaging studies on the online comprehension of noncanonical structures, such as the ERP findings which showed that the semantic P600 effect is observed only when there is a clear alternative meaning that is more plausible than the veridical implausible meaning, and not in other kinds of semantic anomalies that do not have a clear alternative meaning (e.g., Kutas, 1980; Yamada & Neville, 2007). Our results are also consistent with the findings from an ERP study by Jackson, Lorimor, and van Hell (2020), which presented participants with symmetrically plausible passives and found a frontal positivity rather than a posterior P600 at the disambiguating region

(i.e., -ing vs. -ed in the sentence *The policeman was tackling the robber/tackled by the robber...*), suggesting that the comprehension of these symmetrical passives entail a serial revision of thematic roles as each word is processed incrementally. Along the same vein, neuroimaging studies have found that the comprehension of symmetrically plausible passive sentences is supported by the right inferior frontal gyrus, among other regions, which also supports the notion that the processing of noncanonical structures entails thematic role reanalysis (Mack, Meltzer-Asscher, Barbieri, & Thompson, 2013; Meltzer, McArdle, Schafer & Braun, 2010). This process of effortful thematic role revision may recruit cognitive control processes especially when successful completion of the task requires explicitly resolving the syntax-semantics conflict such as in picture matching tasks (e.g., Thothathiri et al., 2018), visual-world paradigms (e.g., Hsu & Novick, 2016), and the thematic role identification task employed in the current study.

Although our findings best support this account of thematic role revision, we cannot rule out the alternative explanation that multiple interpretations may be considered in parallel during incremental processing, but that the non-veridical meaning overrides the veridical meaning only when that alternative meaning is much more compelling. Perhaps the role-reversed meaning is also activated in the processing of symmetrically plausible passives, but this alternative meaning simply does not get selected over the veridical meaning. Follow-up studies are needed to address this question regarding the real-time activation of multiple interpretations across different types of passive sentences (nonreversible, biased, and symmetrical). Additionally, since both the veridical and the role-reversed meanings of the symmetrical sentences were relatively low compared to the plausible interpretations for the nonreversible and biased sentences (see Figure 1), it is possible that the absolute plausibility ratings of the symmetrical sentences were not high enough to induce

competition, and that competition may be observed in symmetrical sentences with higher absolute plausibility levels.

Another alternative explanation is that our findings are in fact consistent with the competing interpretations hypothesis, under the assumption that multiple cues and various constraints are taken into account in a predictive processing architecture, and that multiple interpretations are generated only when the parser encounters strong cues that might require alternative interpretations. When the scenario of the reversed roles has a high event probability and there is a clear thematic and animacy violation, such as in the nonreversible passive sentences (e.g., *The farmer was planted by the corn*), there is a greater conflict between the semantic and syntactic streams that has to be resolved online and may therefore lead to more errors on the thematic role identification questions. The biased passive sentences also denote a violation of semantic/pragmatic plausibility (e.g., *The cat was chased by the mouse*), but not to the same degree as the thematic role anomaly in the nonreversible sentences. For the symmetrically plausible passive sentences (e.g., *The father was consoled by the mother*), hierarchical predictive processing accounts of parsing that assume a commitment to one analysis at a time would not predict any online processing difficulty nor errors on comprehension questions because there are no compelling cues in the linguistic input to trigger a competing interpretation in these sentences. The results from the exploratory analysis on the passive sentences mirrored this pattern, indicating that the presence of a highly plausible alternative interpretation impacted comprehension accuracy for the nonreversible passives, and that the effect was weaker for the biased passives and was completely absent for the symmetrical passives. Since these analyses were exploratory, however, this explanation warrants further investigation in future research.

There are some aspects of the current study's design that may partly explain the absence of the competing interpretations effect in the symmetrical sentences in particular. As shown in Figures 4 and 7, there is high variability in the estimated effect of the role-reversed and relative plausibility predictors for the symmetrical type, especially around the periphery of the plausibility scale (i.e., the lower end of the scale for role-reversed plausibility, and both the lower and higher ends of the scale for relative plausibility). This limited range was an unavoidable consequence of the experimental design, since the symmetrical sentences were designed to be equally plausible. To circumvent this issue, one idea for future research is to investigate the comprehension of passive sentences that are symmetrically implausible (e.g., *The bartender was served by the waiter*) along with passive sentences that are symmetrically plausible (e.g., *The customer was thanked by the clerk*), which would populate data across the entire range of the competing interpretations scale. There are at least two possible outcomes regarding the effect of competing interpretations in the comprehension of these symmetrically implausible sentences. On the one hand, there may be no competition between the two equally implausible interpretations because the role-reversed interpretation is not clearly more plausible than the veridical interpretation. On the other hand, there may be an effect of competing interpretations because the veridical meaning is implausible and would therefore trigger a search for alternative interpretations. Additionally, even though in the current study we did not find an effect of competing interpretations in symmetrically plausible passives, the effect of competing interpretations may be observed in syntactic structures that are more complex than the passive construction, such as sentences with reduced relative clauses or with clefts. Prior work has shown that the misinterpretation effect is smaller in active-passive constructions than in these other types of noncanonical constructions, and that similarity-based

interference between the thematic roles has been observed for other types of noncanonical structures (Gordon, Hendrick, & Johnson, 2001, 2004, 2006; Van Dyke & McElree, 2006).

One notable observation in both experiments is that accuracy performance was similarly high across all conditions (92-98%, across both experiments), restricting the variation in accuracy performance. A possible reason for this high performance is that the task in current study's experiments may have been less cognitively demanding than in Ferreira's (2003) study because on the experimental items, participants were only probed about the agent of the action, and not the patient (also see Dabrowska & Street, 2006). Nevertheless, we did include questions about the patient on 20 of the fillers. Thus in each list, subjects were sometimes asked about the agent as well as occasionally about the patient, and performance was high even on these patient-role questions (91-94% across both experiments). An avenue for future research that may circumvent ceiling performance on passives is to provide a discourse context prior to presentation of the passive sentence, which may be more effective in biasing interpretation towards the schema-consistent but non-veridical interpretation. This expected increase in error rate can provide more variation in comprehension accuracy and can enable a fuller assessment of the role of competing interpretations in the comprehension of passive sentences.

There are some limitations associated with the agent identification task that was used in the current experiments. This comprehension task taps into retrieval of sentence meaning from memory, but the task does not allow an assessment of how competing interpretations may influence comprehension in real time. As discussed earlier, it is possible that there may be competition between more than one interpretation even for the symmetrical passives, but that this effect may not be observed in end-of-sentence question answering accuracy that is measured after

the comprehension processes have been completed. In the present investigation we chose to focus on comprehenders' memory for sentences and how retrieval of the sentence's theta role is influenced by event knowledge, and therefore this study was not designed to address questions regarding the time course of this effect (e.g., whether the observed misinterpretation effects are interpretive or post-interpretive; see Bader & Meng, 2018; Meng & Bader, 2021). A fruitful avenue for future research is to test the competing interpretations hypothesis using measures of moment-by-moment language processing such as eye tracking and electrophysiological recordings. These online measures would provide a stronger test of the serial versus parallel accounts of parsing, and they are also better suited to address whether the processing cost of the role-reversed interpretation is also evident in real-time processing (e.g., longer reading times) in addition to impacting the final interpretation. For instance, one possible mechanism is that the conflict between the syntactic and semantic streams is reconciled during online processing depending on the weight and evidence for the outputs of each of the streams before the final interpretation is reached, and that offline misinterpretations may be attributed to shallow processing and task-specific demands.

Despite the limitations, the current study provides valuable insights into the mechanisms underlying the misinterpretation of passive sentences. Beyond the conventional measure of veridical plausibility, the measures of role-reversed plausibility and relative plausibility enabled an assessment of whether competing interpretations are a viable mechanism underlying the final interpretation of passive sentences. The current study extended prior findings by showing that comprehension accuracy was predicted by not only the plausibility of the veridical interpretation but also the plausibility of the role-reversed interpretation, particularly when this alternative

interpretation is more plausible than the sentence's veridical interpretation. In particular, the novel measure of relative plausibility derived from the slider scale paradigm performed similarly to the measure of absolute plausibility in accounting for comprehension accuracy on the nonreversible and biased sentences, suggesting that this novel measure is a viable method of quantifying the differences in plausibility between two sentence interpretations. The present study also extends prior investigations by providing insight into the comprehension of symmetrically plausible passives, which have been less studied than passive sentences that contain a syntax-semantics conflict. These symmetrical sentences typically convey a reciprocal action (e.g., hugging, kissing, greeting) or an action that both entities are equally likely to do to each other (e.g., thanking, calling, congratulating). The fact that these symmetrical passives contain two equally compelling sets of thematic roles is a unique feature that can contribute to the debate regarding parallel versus serial processes underlying thematic role assignment in noncanonical structures.

In sum, the current study took a step towards testing the hypothesis that the misinterpretation of passive sentences arises due to competition between the veridical interpretation and a non-veridical, role-reversed interpretation that are both considered in parallel. The findings from both experiments indicated that question answering accuracy decreased as the plausibility of the role-reversed meaning increased. However, this negative relationship was observed only in sentences that denote a syntax-semantics conflict (i.e., the nonreversible sentences and biased sentences). Crucially, question answering accuracy was not predicted by the plausibility of the role-reversed meaning in sentences that have two equally compelling interpretations (i.e., symmetrically plausible sentences). This suggests that the misinterpretation of passive sentences is not driven by parallel competition between the two meanings, but rather that plausibility-based

heuristics exert influence via downstream thematic role revision of an implausible veridical interpretation when schematic knowledge suggests a more plausible, non-veridical meaning. Competing linguistic interpretations may be generated only when there is a reason in the input or the comprehension task to do so. Further work is required, however, to determine whether the alternative, role-reversed meaning is activated and considered during real-time processing of passive sentences.

References

- Amichetti, N. M., White, A. G., & Wingfield, A. (2016). Multiple solutions to the same problem: utilization of plausibility and syntax in sentence comprehension by older adults with impaired hearing. *Frontiers in Psychology, 7*, 789.
- Bader, M., & Meng, M. (2018). The misinterpretation of noncanonical sentences revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(8), 1286.
- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews, 59*(1), 55-73.
- Christianson, K., Luke, S. G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(2), 538.
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*(4), 368-407.
- Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes, 42*(2), 205-238.
- Dabrowska, E., & Street, J. (2006). Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences, 28*(6), 604-615.
- Ferreira, F., & Patson, N. D. (2007). The ‘good enough’ approach to language comprehension. *Language and Linguistics Compass, 1*(1-2), 71-83.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*(2), 164-203.

- Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11-15.
- Fodor, J. D., & Inoue, A. (1998). Attach anyway. In *Reanalysis in sentence processing* (pp. 101-141). Springer, Dordrecht.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6(4), 291-325.
- Gernsbacher, M. A. (1991). Cognitive processes and mechanisms in language comprehension: The structure building framework. *Psychology of Learning and Motivation*, 27, 217-263.
- Gibson, E., & Pearlmutter, N. J. (2000). Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research*, 29(2), 231-240.
- Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, 110(20), 8051-8056.
- Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(6), 1411.
- Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, 51(1), 97-114.
- Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of experimental psychology: Learning, Memory, and Cognition*, 32(6), 1304.
- Hagoort, P., Baggio, G., and Willems, R. (2009). "Semantic unification," in *The New Cognitive Neuroscience*, ed. M. S. Gazzaniga (Cambridge, MA: MIT Press), 56, 1-18.

- Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 32(2), 101-123.
- Hestvik, A. (1995). Reflexives and ellipsis. *Natural Language Semantics*, 3(2), 211-237.
- Hsu, N. S., & Novick, J. M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychological Science*, 27(4), 572-582.
- Jackson, C. N., Lorimor, H., & van Hell, J. G. (2020). The ERP correlates of thematic role assignment for passive versus active sentences. *Journal of Neurolinguistics*, 54, 100886.
- Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, 69(5), 1013-1040.
- 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.
- Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363.
- Kolk, H. H., Chwilla, D. J., Van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, 85(1), 1-36.
- Kos, M., Vosse, T. G., Van Den Brink, D., & Hagoort, P. (2010). About edible restaurants: conflicts between syntax and semantics as revealed by ERPs. *Frontiers in Psychology*, 1, 222.
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23-49.

- Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*, 31(5), 602-616.
- Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17(1), 117-129.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203-205.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375-419.
- Lewis, R. L. (2000). Falsifying serial and parallel parsing models: Empirical conundrums and an overlooked paradigm. *Journal of Psycholinguistic Research*, 29(2), 241-248.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological review*, 101(4), 676.
- Mack, J. E., Meltzer-Asscher, A., Barbieri, E., & Thompson, C. K. (2013). Neural correlates of processing passive sentences. *Brain Sciences*, 3(3), 1198-1214.
- McRae, K., & Ferretti and Liane Amyote, T. R. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, 12(2-3), 137-176.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38(3), 283-312.

- Meltzer, J. A., McArdle, J. J., Schafer, R. J., & Braun, A. R. (2010). Neural aspects of sentence comprehension: syntactic complexity, reversibility, and reanalysis. *Cerebral Cortex*, 20(8), 1853-1864.
- Meng, M., & Bader, M. (2021). Does comprehension (sometimes) go wrong for noncanonical sentences?. *Quarterly Journal of Experimental Psychology*, 74(1), 1-28.
- Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience*, 22(12), 2886-2898.
- Paolazzi, C. L., Grillo, N., Alexiadou, A., & Santi, A. (2019). Passives are not hard to interpret but hard to remember: evidence from online and offline studies. *Language, Cognition and Neuroscience*, 34(8), 991-1015.
- Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57(3), 348-379.
- Sanford, A. J. (2002). Context, attention and depth of processing during interpretation. *Mind & Language*, 17(1-2), 188-206.
- Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, 69(2), 104-120.
- Slobin, D. I. (1966). Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal learning and Verbal Behavior*, 5(3), 219-227.
- Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, 105(2), 477-488.

- Thothathiri, M., Asaro, C. T., Hsu, N. S., & Novick, J. M. (2018). Who did what? A causal role for cognitive control in thematic role assignment during sentence comprehension. *Cognition*, 178, 162-177.
- Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55(2), 157-166.
- Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In *Reading as a perceptual process* (pp. 621-648). North-Holland.
- Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, 45(2), 225-258.
- Van Herten, M., Chwilla, D. J., & Kolk, H. H. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, 18(7), 1181-1197.
- Van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22(2), 241-255.
- Yamada, Y., & Neville, H. J. (2007). An ERP study of syntactic processing in English and nonsense sentences. *Brain Research*, 1130, 167-180.

Appendix

Experimental items for both Experiments 1 and 2. Only the items in the Active-Plausible condition are shown (for the symmetrical sentences this corresponds to Active-Order1). The implausible condition (Order2 for the symmetrical sentences) was created by reversing the order of the two nouns, such as *The apron wore the chef*. The passive condition was the passivized version of each sentence, such as *The apron was worn by the chef*.

Nonreversible sentences:

1. The chef wore the apron.
2. The farmer planted the corn.
3. The mouse ate the cheese.
4. The dog buried the bone.
5. The editor reviewed the paper.
6. The plumber fixed the drain.
7. The runner won the race.
8. The cow chewed the cud.
9. The DJ played the music.
10. The secretary typed the letter.
11. The artist painted the picture.
12. The termite chewed the wood.
13. The chicken laid the egg.
14. The ant built the hill.
15. The doctor took the X-ray.
16. The pirate buried the treasure.
17. The nurse gave the shot.
18. The dentist pulled the tooth.
19. The beaver gnawed the tree.
20. The bulldozer pushed the dirt.
21. The tailor hemmed the skirt.
22. The pilot flew the plane.
23. The chauffeur drove the car.
24. The conductor led the orchestra.
25. The geologist examined the fossil.
26. The janitor mopped the floor.
27. The shopper bought the dress.
28. The bully stole the cookie.

29. The professor graded the exam.
30. The priest delivered the sermon.
31. The designer decorated the room.
32. The cook chopped the onion.
33. The knight invaded the castle.
34. The journalist wrote the article.
35. The barista brewed the coffee.
36. The architect designed the house.
37. The busboy cleaned the table.
38. The musician composed the song.
39. The mechanic repaired the truck.
40. The gardener tended the plant.
41. The athlete lifted the barbell.
42. The firefighter extinguished the flame.
43. The neighbor heard the noise.
44. The sorcerer cast the spell.
45. The traveler photographed the scenery.
46. The mobster robbed the museum.
47. The hiker explored the mountain.
48. The contractor repaired the roof.

Biased sentences:

1. The dog bit the man.
2. The cook ruined the food.
3. The bird ate the worm.
4. The cat chased the mouse.
5. The soldier protected the villager.
6. The lawyer sued the doctor.
7. The teacher quizzed the student.
8. The cop pursued the thief.
9. The waitress served the man.
10. The owner fed the cat.
11. The detective investigated the suspect.
12. The doctor treated the patient.
13. The politician deceived the voter.
14. The hiker killed the mosquito.
15. The horse threw the rider.
16. The golfer hit the ball.
17. The hunter shot the deer.
18. The frog ate the fly.

19. The ghost scared the boy.
20. The angler caught the fish.
21. The matador dodged the bull.
22. The officer arrested the citizen.
23. The paramedic carried the child.
24. The counselor advised the student.
25. The parent scolded the toddler.
26. The snake devoured the rabbit.
27. The judge sentenced the terrorist.
28. The babysitter watched the child.
29. The lecturer instructed the class.
30. The supervisor hired the intern.
31. The cop guarded the mayor.
32. The scout discovered the model.
33. The victim accused the perpetrator.
34. The audience applauded the musician.
35. The ranger guided the tourist.
36. The lifeguard saved the surfer.
37. The firefighter rescued the family.
38. The mother tickled the baby.
39. The cameraman filmed the president.
40. The magician entertained the crowd.
41. The principal reprimanded the student.
42. The host awarded the recipient.
43. The senator bribed the press.
44. The carpenter fired the apprentice.
45. The bodyguard escorted the prince.
46. The local welcomed the visitor.
47. The commander led the troop.
48. The lawyer defended the client.

Symmetrical sentences:

1. The boy kicked the girl.
2. The girlfriend kissed the boyfriend.
3. The sister hugged the brother.
4. The committee introduced the chairman.
5. The runner saw the driver.
6. The woman called the girl.
7. The man visited the woman.
8. The boy touched the man.

9. The producer recognized the director.
10. The priest approached the rabbi.
11. The child loved the puppy.
12. The butcher despised the baker.
13. The team chose the player.
14. The clerk thanked the customer.
15. The teacher greeted the parent.
16. The mother adored the son.
17. The realtor faxed the buyer.
18. The catcher signaled the pitcher.
19. The broker phoned the client.
20. The guest insulted the host.
21. The model met the photographer.
22. The prime minister embraced the pope.
23. The witch praised the wizard.
24. The applicant contacted the company.
25. The father consoled the mother.
26. The gangster killed the sheriff.
27. The accountant helped the salesman.
28. The boy teased the girl.
29. The editor consulted the author.
30. The governor invited the diplomat.
31. The policeman respected the astronaut.
32. The brother pestered the sister.
33. The scientist honored the doctor.
34. The guard defeated the intruder.
35. The player tackled the opponent.
36. The foreman assisted the worker.
37. The zombie kicked the vampire.
38. The trooper confronted the shooter.
39. The husband accused the wife.
40. The inspector shot the spy.
41. The artist mocked the reviewer.
42. The tiger attacked the leopard.
43. The prankster deceived the magician.
44. The banker paid the cashier.
45. The senator supported the campaign.
46. The actress congratulated the actor.
47. The waiter served the bartender.
48. The bear hunted the lion.

CHAPTER 3

Shallow processing in the comprehension of elliptical verb phrases

Abstract

Previous research has shown that comprehenders systematically misinterpret implausible sentences that require a noncanonical order of thematic role assignment, suggesting that a plausibility strategy can lead to a non-veridical representation of the linguistic input. Compared to sentences in which the linguistic content is overtly stated, relatively little is known about how plausibility heuristics influence the final interpretation of sentences that contain elided material. This experiment investigates the factors that influence the accuracy at which interpretations are assigned to verb-phrase ellipses with passive antecedents that varied in semantic plausibility. Extrapolating the competing interpretations hypothesis proposed by Chantavarin (2021) from passive sentences to elliptical clauses, we expect that the language comprehension system will be more likely to select an alternative, competing interpretation of the elided content when that interpretation is highly supported by the discourse context. We also assess the possibility that shallow processing heuristics are more likely to exert influence in the comprehension of elided content, compared to the same sentence without ellipsis.

Keywords: sentence comprehension, thematic role assignment, misinterpretation errors, verb-phrase ellipsis, event plausibility, shallow processing

1. Introduction

A fundamental component of language comprehension is syntactic parsing, which establishes the dependencies between the words in a sentence to build a coherent representation of the sentence's meaning. Standard models of parsing assume that sentence interpretation is generated through a set of algorithms that organizes the words into a syntactic structure (e.g., Frazier & Fodor, 1978; MacDonald, Pearlmutter, & Seidenberg, 1994; Gibson, 2000), and this syntactic representation is then used to construct a semantic representation of the intended meaning. However, a number of subsequent studies have shown that the final interpretation of a sentence may not always be consistent with the surface details of the linguistic representation generated via syntactic analysis. Such misinterpretation effects have been shown for garden-path sentences (Christianson, Williams, Zacks, & Ferreira, 2006; Christianson, Hollingworth, Halliwell, & Ferreira, 2001) as well as for unambiguous, noncanonical structures such as passive sentences (Ferreira, 2003; Christianson, Luke, & Ferreira, 2010). For instance, people make more errors identifying the sentence's thematic roles in passive sentences than active sentences, and people also made more errors in implausible sentences than plausible sentences, suggesting that heuristic strategies can distort sentence interpretation (Ferreira, 2003). These findings are accounted for by more recent approaches to parsing such as the good-enough model, which proposes that language processing involves generating a "good-enough" representation for the task at hand, rather than delivering a detailed, accurate representation (Ferreira & Patson, 2007; Ferreira, Bailey, & Ferraro, 2002).

Besides passive sentences, ellipsis is another noncanonical structure that may be subject to misinterpretation effects. Ellipsis is a type of anaphoric expression in which the surface form of the

sentence contains missing (or elided) linguistic material, yet the sentence interpretation can be recovered from the local (usually preceding) context. Examples of ellipsis include noun-phrase (NP) ellipsis (1a), verb-phrase (VP) ellipsis (1b), and sluicing (1c). The current study focuses on VP ellipses.

- (1) a. The judges liked Vera's artwork and they also liked Dan's [].
b. My mom thought the movie was great but my sister didn't [].
c. She said something bad about me, but I don't know what [].

Successful interpretation of ellipses requires retrieving the syntax and meaning of the relevant linguistic material (i.e., the antecedent) to fill out the unpronounced material at the ellipsis site, denoted as brackets in example (1) above. Most studies on the processing of ellipses have investigated questions that pertain to the detailed syntactic analysis of elliptical structures, such as the linguistic properties of the antecedent that are relevant for ellipsis resolution, and the nature of the linguistic content represented at the ellipsis site (for comprehensive reviews, see Phillips & Parker, 2014 and Merchant, 2019). Beyond this scope, ellipses can also provide valuable insight into how plausibility heuristics influence the comprehension of noncanonical sentences. Unlike fully specified, non-elided structures, the linguistically correct interpretation of elliptical clauses crucially depends on retrieving the antecedent's surface form, and not just the gist of the semantic content. For instance, to understand a sentence containing a passive-voice elided VP such as in example (2) below, comprehenders must retrieve the precise form of the passive antecedent *scolded by the teacher* to assign the correct interpretation to the elided material (i.e., that the friend was scolded by the teacher too). If comprehenders retrieve a semantically equivalent but syntactically

different version of the antecedent (e.g., *The teacher scolded the student*), this would lead to a wrong interpretation of the elided constituent (i.e., that the friend was scolding the student too).

(2) The student was scolded by the teacher, and the friend was [] too.

Past research on misinterpretation effects has mostly investigated shallow processing in overtly stated content (e.g., passive sentences), but relatively little is known about the factors that impact the accuracy of interpreting elliptical content. To make progress in understanding how plausibility heuristics influence the processing of ellipses, the current study addresses two aspects regarding the comprehension of passive-voice ellipses. First, how accurately are passive-voice antecedents processed in memory, and how does this impact the accuracy of assigning an interpretation to a subsequent elliptical clause? Second, are passive-voice ellipses more prone to shallow processing than their corresponding non-elided structure (i.e., non-elided passive sentences)? To set the stage for this study, in the following sections we provide a brief overview on the misinterpretation of passive sentences, followed by a discussion of the literature on the processing of ellipses.

1.1 The processing of passive sentences

In English sentences, the agent of the action is typically mentioned before the patient/theme (or recipient) of the action, such as in example (3a) below. In contrast, passive sentences like the example in (3b) are noncanonical because this thematic order is reversed, and thus, during incremental processing, the thematic roles have to be assigned in a nonstandard order.

(3) a. The teacher_{AGENT} scolded the student_{PATIENT}.

b. The student_{PATIENT} was scolded by the teacher_{AGENT}.

As such, misinterpretation effects have been consistently shown in the comprehension of passive sentences. Initial evidence for this was reported in a study by Ferreira (2003) that investigated the influence of heuristic, good-enough processing on the comprehension of passive sentences and object-clefts, both of which required the patient to be assigned before the agent. Analyses on question answering accuracy indicated that passives and object-clefts were systematically misinterpreted, suggesting a reliance on the agent-before-patient heuristic in parsing (for a competing view, see Paolazzi, Grillo, Alexiadou, & Santi, 2019). The accuracy of identifying thematic roles was also lower when the sentence conveyed an implausible meaning, suggesting a reliance on plausibility heuristics as well. Subsequent studies have found similar misinterpretation effects in passive sentences (Christianson et al., 2010; Gibson, Bergen, & Piantadosi, 2013; Dabrowska & Street, 2006; Amichetti, White, & Wingfield, 2016). Findings from event-related potential (ERP) studies have also shown that semantic heuristics can sometimes override algorithmic parsing, especially when the sentence contains conflicting syntactic and semantic information (Kim & Osterhout, 2005; Van Herten, Kolk, & Chwilla, 2005; Hoeks, Stowe, & Doedens, 2004).

These findings suggest that the language comprehension system may rely on the syntactic stream as well as a potentially independent semantic memory-based stream in parsing (Kuperberg, 2007; but see Kuperberg, 2016). If the parser relies on plausibility heuristics in assigning thematic roles, how are the output from the syntactic algorithm and from semantic heuristics reconciled? One possibility suggested by Chantavarin (2021) is that the parser considers both the veridical interpretation as well as other alternative interpretations that are also compelling, and the

likelihood of misinterpretation depends on the degree to which the alternative interpretation was plausible (or relatively more plausible than the veridical interpretation). This so-called *competing interpretations hypothesis* predicts that the greater the plausibility of the competing interpretation, the more errors people should make on comprehension questions (i.e., more misinterpretations). In two comprehension studies, Chantavarin (2021) found that the competing interpretation's role-reversed plausibility and relative plausibility accounted for question answering accuracy on passive sentences that denoted a syntax-semantics conflict (e.g., *The secretary was typed by the letter; The cop was guarded by the mayor*), but not on sentences whose argument orders yielded two equally plausible meanings (e.g., *The husband was accused by the wife*). These results provided preliminary evidence against the competing interpretations account which assumed concurrent activation of all possible analyses that compete for selection, and instead the findings are more compatible with the view that misinterpretations reflect thematic role revision of the initially adopted interpretation when its meaning contradicts real-world plausibility.

However, from this study alone we cannot completely rule out the competing interpretations hypothesis. For instance, it is possible that the plausibility of the competing interpretation may influence the final interpretation if the passive sentences were presented within a discourse context that provides a stronger bias towards a schema-consistent but incorrect interpretation, and if the noncanonical structure being tested was more prone to misinterpretations. To explore this possibility, in the present study we extend the assessment of the competing interpretations hypothesis from isolated passive sentences to short passages containing passive-voice ellipses, another movement-derived noncanonical structure whose interpretation may also be systematically distorted due to a reliance on a plausibility strategy.

1.2 *The processing of ellipses*

Ellipsis has been a widely investigated topic in theoretical linguistics, but recent years have also seen a growing interest in experimental studies on the processing of ellipses (e.g., Parker, 2018; Frazier & Duff, 2019; Kim & Runner, 2018; Kim, Brehm, & Yoshida, 2019; Paape, Hemforth, & Vasishth, 2018; Xiang, Grove, & Merchant, 2019). Empirical studies on the processing of ellipses have centered around two main debates (for further discussion see Phillips & Parker, 2014). The first debate pertains to the nature of the antecedent that is relevant for ellipsis resolution, specifically, whether the interpretation of ellipses involves the recovery of an antecedent that is syntactically or semantically identical to the ellipsis site. Most of the relevant empirical evidence comes from acceptability judgment studies, and they support both the syntactic identity account (Sag, 1976; Lappin, 1992) as well as the semantic identity account (Dalrymple, Shieber, & Pereira, 1991; Hardt, 1999). There is a hybrid view which has received support as well. According to this view, how ellipsis resolution is achieved may depend on processing constraints such as discourse coherence (Kehler, 2000) and/or information structure (Kertz, 2013), which determine whether a syntactically or semantically matching antecedent is required to interpret the ellipsis. The second prominent debate is whether there is a detailed syntactic representation (or null copy) of the antecedent at the ellipsis site. This question also remains inconclusive because the results have been mixed across different studies (e.g., Kaan, Wijnen, Swaab, 2004; Frazier & Clifton, 2000; Martin & McElree, 2008; for further discussion, see Phillips & Parker, 2014).

1.2.1 *Shallow processing of ellipses*

As discussed in the preceding section on the processing of passive sentences, the evidence for misinterpretations suggests that comprehenders may not always build a detailed, accurate

representation of the linguistic input (Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007; Sanford & Sturt, 2002). Several studies have investigated shallow processing (specifically, the reliance on the plausibility strategy) in overtly stated (or non-elided) linguistic input, but relatively little is known about shallow processing of elided content. The interpretation of ellipses may not depend solely on the linguistic representation of the antecedent because the processing of anaphora has been argued to require the retrieval of not only the surface features of the preceding text, but also the conceptual and semantic information in the discourse or mental model (Dalrymple, Shieber, & Pereira, 1991; Hardt, 1993; Hestvik, 1995; for detailed discussions see Garnham & Oakhill, 1990 and Garnham, 2001). Thus, we might expect that the process of assigning an interpretation to ellipses may also be subject to biases and distortions from the influence of contextual and world knowledge, akin to the misinterpretation effects that have been shown for other noncanonical structures that do not contain elided content.

Scant research has been done on this topic, with the exception of a few studies. In a study that directly probed the accuracy at which comprehenders assign interpretations to elided constituents, Garnham and Oakhill (1987) investigated whether people's final interpretation of elliptical VPs can be systematically biased by real-world plausibility. In a self-paced reading paradigm, participants read three-sentence passages containing a filler sentence introducing the setting (*It was a busy morning at the hospital*), followed by a plausible passive sentence denoting a biased, plausible transitive event (e.g., *The patient had been examined by the doctor*), and then an elliptical VP that was either plausible (*The child had too*) or implausible (*The nurse had too*). In addition, the distance between the ellipsis and the passive antecedent was manipulated by including an intervening clause in the antecedent (*The patient had been examined by the doctor /during*

the ward round/). Participants responded to comprehension questions targeting thematic role assignment (e.g., *Did the doctor examine the nurse? / Did the nurse examine the patient?*). Garnham and Oakhill found that question answering accuracy was lower on implausible ellipses than plausible ones, indicating that people tended to select a more plausible, schema-consistent interpretation (e.g., that doctors and nurses are likely to examine patients and children) over the veridical but implausible interpretation (e.g., that doctors examine nurses), even if the adopted interpretation is outright incompatible with the syntactic structure. This study provided evidence that shallow processing also occurs in elided structures, particularly when the elided content denotes an implausible event.

Garnham and Oakhill (1987) also found longer clause reading times and question answering times for implausible ellipses than plausible ones, and these longer latencies were also observed as the distance increased between the antecedent and the ellipsis. Based on these findings, they concluded that people tried to interpret text accurately, but it became more difficult when there is a competing, more plausible interpretation suggested by the context. This notion of competition or interference between the veridical interpretation and an alternative interpretation in ellipsis resolution appears compatible with the competing interpretations hypothesis proposed by Chantavarin (2021). However, there are two limitations about this claim that we intend to address in the present study. First, it is difficult to definitively conclude that Garnham and Oakhill's (1987) results from the clause reading times indicate a competition mechanism because this measure could either reflect competing representations in a parallel-competitive parsing architecture as proposed by Chantavarin (2021; also see McRae, Spivey-Knowlton, & Tanenhaus, 1998; McRae, Spivey & Tanenhaus, 1998), or alternatively, it could reflect the cost of thematic

analysis in a two-stage parsing architecture (Frazier & Fodor, 1978; Fodor & Inoue, 2000; Van Gompel, Pickering, & Traxler, 2000). To adjudicate between these two possibilities, a more direct test is needed to establish whether the parallel-competition account can explain the misinterpretation of ellipses. Second, if there is indeed competition between the veridical interpretation and an alternative, non-veridical interpretation suggested by the context, it is unclear whether this competition occurs only when the alternative interpretation is clearly more plausible, or whether competition also occurs when the sentence has two equally plausible interpretations. There may be competition or interference between the veridical interpretation and an alternative interpretation when the antecedent conveys a symmetrically plausible event or a reciprocal action, according to the competing interpretations hypothesis which predicted that shallow processing may exert influence even in symmetrically plausible sentences. Although Chantavarin (2021) did not find evidence for this when they assessed the comprehension of isolated passive sentences, this possibility may be observed in the processing of elided structures presented with a preceding discourse context that sets up a schematic expectation, which may increase the reliance on schematic knowledge during parsing. Thus, the present study aims to directly assess the competing interpretations hypothesis in the comprehension of ellipses by explicitly quantifying the plausibility of the alternative, non-veridical interpretation relative to the plausibility of the veridical interpretation, and to determine whether this measure accounts for comprehension accuracy.

There is some evidence that the processing of ellipses may involve concurrent activation of competing analyses. In a series of experiments, Shapiro and colleagues (1995; 2003) used a cross-modal lexical decision task to investigate whether competing interpretations are activated during

the processing of elliptical VPs, although their study was designed to characterize the time-course of activated interpretations for elided constituents rather than the accuracy of the final interpretation. They used elided VPs that contained overt anaphors (e.g., pronouns) such as *The policeman perjured himself, and the fireman did too*. In this example, there are two linguistically licensed interpretations of the elided constituent in the second clause (*the fireman did too*). On the one hand, the elided content can mean that *The fireman perjured the policeman*, an interpretation in which the copied reflexive pronoun (*himself*) refers to the same referent as the subject of the first clause (*policeman*). This so-called “strict” reading is implausible and perhaps also ungrammatical because a person cannot perjure someone else. On the other hand, the elided material can mean that *The fireman perjured himself*, an interpretation in which the copied pronoun is coindexed with the subject of the second clause (*fireman*) instead. This so-called “sloppy” reading is more plausible and is the final interpretation that people prefer (Frazier & Clifton, 2000). Results from the lexical priming task showed facilitated lexical decision times to words that were related to the subject noun phrase in the first clause (*policeman*) and in the second clause (*fireman*). This suggests that both of these nouns were reactivated at the ellipsis site (*...did [] too*) and that comprehenders activated both the strict reading and the sloppy reading of the elided VP during online processing, even though the latter is greatly preferred as the final interpretation. These results suggest that when there is more than one viable interpretation of an elliptical phrase, those interpretations are computed in parallel during online processing.

A second research question addressed in the current study is whether the degree to which comprehenders rely on the plausibility strategy when processing passive sentences varies depending on whether the VP is elided versus non-elided. We might expect that shallow processing exerts

greater influence in passive-voice ellipsis than in non-elided passive-voice sentences because the comprehension of ellipsis requires retrieval of the antecedent from memory to construct an interpretation at the ellipsis site, which may make the parser more susceptible to shallow processing. Specifically, the processing of elided structures may recruit a greater reliance on schematic knowledge over the antecedent's veridical structure and meaning.

1.3 Current study

This study extends previous findings on the shallow processing of passive sentences to elided, passive-voice VPs. The goals of this study are twofold. The first is to verify the robustness of the competing interpretations hypothesis proposed by Chantavarin (2021) in accounting for the misinterpretation of ellipses. Specifically, we ask whether the accuracy of ellipsis interpretation depends on the extent to which the alternative interpretation is plausible (or, more plausible than the veridical interpretation). The second goal is to compare the degree of shallow processing in elided passive VPs compared to non-elided passive sentences. Specifically, we ask whether passive sentences are more likely to be misinterpreted when a portion of the sentence (here, the VP) is elided, compared to when the entire linguistic content is overtly stated. To these aims, we used the same passive sentences as in Chantavarin's (2021) study, with two modifications. For each item we added a preceding filler sentence that introduces the setting prior to the passive sentence, and we added a plausible elliptical VP after the passive sentence, resulting in a three-sentence passage (see Table 1 for examples). Following Chantavarin (2021), the passive antecedents were categorized into 3 types (nonreversible, biased, and symmetrical), according to the degree of meaning bias between the veridical interpretation and the alternative interpretation denoted by the reversed set of thematic roles. We did not include the active sentences because we are interested specifically in

the misinterpretation of passive sentences. As in Chantavarin (2021), we collected plausibility ratings to quantify the plausibility of the veridical interpretation as well as the role-reversed interpretation of the passive sentences. We will assess whether accuracy on comprehension questions probing thematic role identification in the elliptical clause depends on the plausibility of the role-reversed interpretation of the antecedent.

1.3.1 Hypotheses and predictions

Our first hypothesis is that the competing interpretations hypothesis accounts for the interpretation of passive-voice elliptical VPs (see Figure 1 for a schematic diagram). If this is the case, then the accuracy of ellipsis interpretation should depend on the plausibility of the antecedent's alternative, non-veridical meaning. Given a supporting discourse context and an elliptical structure that may be more prone to misinterpretations, and as predicted by the competing interpretations hypothesis, this effect should be observed for all sentence types, even for the symmetrical passives. An alternative outcome is that there is no relationship between the plausibility of the alternative meaning and question answering accuracy for the symmetrically plausible passives. This would replicate the findings from Chantavarin (2021) and provide further supporting evidence that there is no parallel competition between the two equally plausible meanings.

The second hypothesis is that having to interpret passive-voice elliptical clauses may allow semantic heuristics to exert greater influence than when people have to interpret non-elided passive sentences. More comprehension errors may occur in the process of retrieving the representation for the passive antecedent to fill out the missing portion of the elliptical clause. If this hypothesis is true, we should observe a greater tendency to misinterpret passives when that

representation is retrieved to interpret an elided VP (as in the current study), compared to when the passives are presented in isolation (as in Chantavarin, 2021). This increased rate of misinterpretation should be observed specifically in the implausible passives and symmetrical passives because they have a potentially competing role-reversed interpretation. On the other hand, if elided content is not more prone to shallow processing than overtly stated content, then the rate of misinterpretation should be similar whether comprehension is probed on passive elliptical clauses or when the passives are presented in isolation.

Table 1

Experimental design and conditions

Sentence type	Argument Order in Critical Sentence	
	Plausible (Order1)	Implausible (Order2)
Nonreversible	Everyone at the office was preparing for the meeting. The letter was typed by the secretary . The memo was too.	Everyone at the office was preparing for the meeting. The secretary was typed by the letter . The memo was too.
Biased	It had been a busy morning at the hospital. The patient was treated by the doctor . The child was too.	It had been a busy morning at the hospital. The doctor was treated by the patient . The child was too.
Symmetrical	The store was filled with people shopping for the holidays. The clerk was thanked by the customer . The manager was too.	The store was filled with people shopping for the holidays. The customer was thanked by the clerk . The manager was too.

Note. The noun phrases in the passive sentences are bolded here for illustrative purposes only.

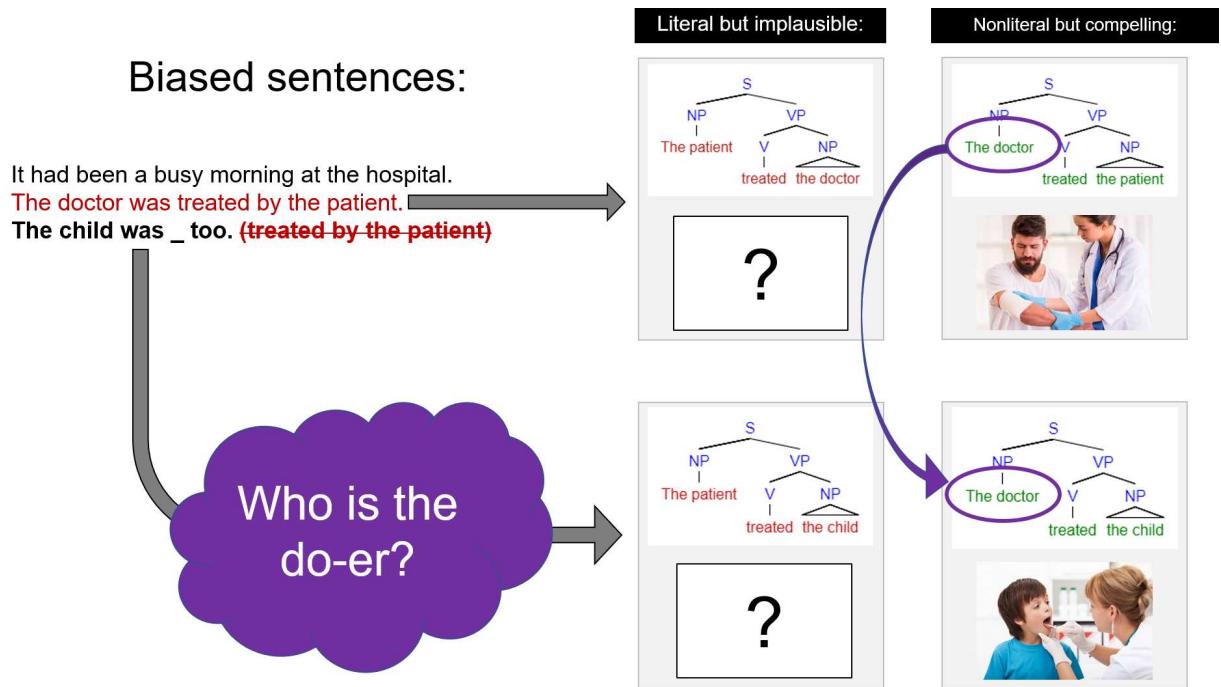


Figure 1. Schematic diagram illustrating the hypothesized effect of competing interpretations on the accuracy of ellipsis interpretation. The figure shows an example item from the Biased-Implausible condition (*The doctor was treated by the patient*). Accuracy on the agent identification question (“Who is the do-er?”) should be lower as the plausibility of the role-reversed interpretation increases, reflecting the parser’s tendency to select a more plausible, role-reversed interpretation of the antecedent (*The doctor treated the patient*) to reconstruct an interpretation in the elided portion of the elliptical clause (*The child was treated by the doctor too*).

2. Absolute plausibility norming study

In this norming study, we collected the absolute plausibility ratings on the experimental items that will be used as a predictor of comprehension accuracy in the main experiment. The experimental stimuli consisted of 126 passive sentences selected from Chantavarin’s (2021) study.

There were three sentence types as follows. A third of the items were nonreversible because one of the arguments was an inanimate noun (e.g., *the apron was worn by the chef*). Another third of the items were biased, in which one argument order was more plausible than the reversed order (e.g., *the man was bitten by the dog*). The last third were symmetrical items in which the two argument orders were designed to be equally plausible (e.g., *the clerk was thanked by the customer*). A filler sentence was created for each item to introduce the general setting of the event denoted in the experimental sentence, as well as an elliptical VP that contained a noun phrase that was a plausible recipient of the action (see Table 1 for examples, and the Appendix for the full list of stimuli). The agent and patient of the action in the plausible/order1 version of the experimental sentence are henceforth labelled as NP1 and NP2, respectively, whereas the subject NP in the elliptical VP is labelled as NP3 (see Figure 3 for an illustration).

We assessed the plausibility of the two argument orders for each experimental sentence (NP1-NP2 and NP2-NP1), as well as the plausibility of NP1 and NP2 as agents of the elided phrase (NP1-NP3 and NP2-NP3). For completeness, we also assessed the plausibility of NP3 as an agent of the two noun phrases from the experimental sentence (NP3-NP1 and NP3-NP2). In this plausibility rating task, the experimental sentences were presented in the active voice to avoid sentence misinterpretations, and each item was presented with its corresponding filler (or context) sentence to assess the plausibility of the denoted event in that particular discourse context. Ninety participants took part in this norming study. The 126 experimental items were counterbalanced across six lists so that participants saw only one order of the noun phrases for each item (i.e., NP1-NP2, NP2-NP3, NP1-NP3, and their reversed orders). Participants were asked to rate how semantically plausible each sentence was from 1 (Very Implausible) to 7 (Very Plausible), and they

were presented with five practice trials describing the scale to ensure that they understood the plausibility rating task.

The mean plausibility ratings across all three sentence types are presented in Figure 4 and Table 2. The absolute plausibility ratings for the critical sentence closely replicated the means reported by Chantavarin (2021). For the nonreversible items and biased items, the plausible argument order of the critical sentence (NP1-NP2) received a much higher plausibility rating than the implausible order (NP2-NP1), indicating that these critical sentences had the appropriate semantic properties. For the symmetrical items, the two argument orders of the critical sentence (NP1-NP2 and NP2-NP1) were rated similarly high, affirming that these sentences were symmetrically plausible as designed. These absolute plausibility ratings will be used as continuous predictors in the analyses of the main experiment. In addition to quantifying the semantic properties of the critical sentences, a second aim of this norming study was to assess the semantic properties of the sentences involving the noun phrase in the elliptical clause (NP3) as well. Crucially, the NP1-NP3 order was rated as highly plausible across all three sentence types, confirming that the noun phrase in the elliptical VP (NP3) was a plausible patient of the action with NP1 as the agent. These ratings for NP1-NP3 were similarly high to the ratings for the plausible/order1 version of the critical sentence (NP1-NP2), indicating that both NP2 and NP3 are plausible patients of NP1.

The remaining comparisons between the noun phrase combinations were of secondary interest. Unsurprisingly, for the nonreversible and biased sentences, NP2 was not a highly plausible agent of NP3, and NP3 was not a highly plausible agent of neither NP1 nor NP2. For the symmetrical sentences, the plausibility ratings across all six NP combination orders were rated

similarly high, indicating that NP1, NP2, and NP3 are all symmetrically plausible agents and patients of the verb.

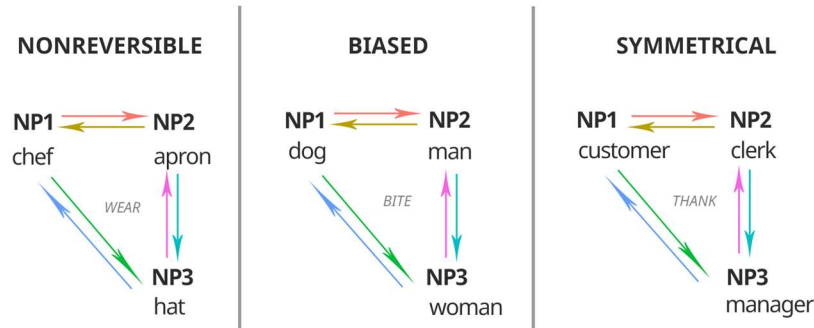


Figure 3. Schematic diagram of the relationship between the three critical noun phrases for each sentence type. NP1 and NP2 refer to the agent and the patient of the plausible/order1 version of the critical sentence, respectively (e.g., *The apron_{NP2} was worn by the chef_{NP1}*). NP3 refers to the noun phrase in the elliptical VP (e.g., *The hat_{NP3} was too*). Arrows denote the six possible thematic relationships between each pairwise combination of the three NPs.

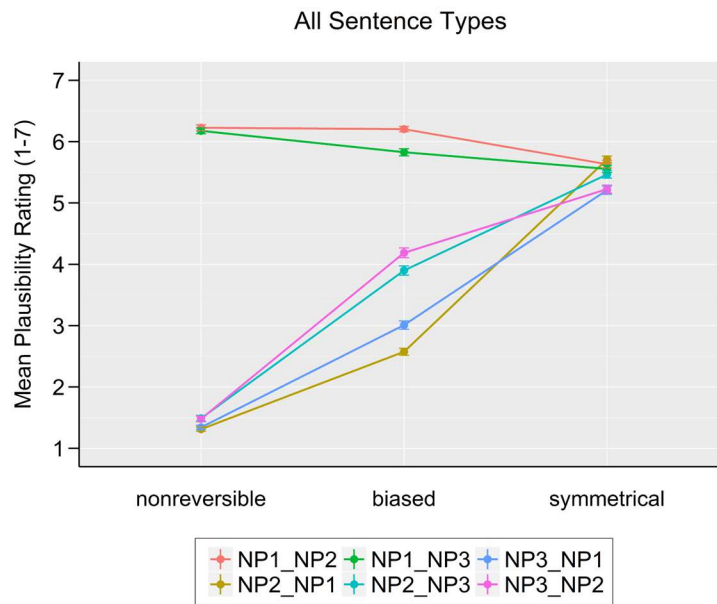


Figure 4. Mean absolute plausibility ratings for the 6 noun-phrase order combinations across the nonreversible, biased, and symmetrical sentence types.

Table 2

Absolute plausibility ratings for the 6 noun phrase combinations across the nonreversible, biased, and symmetrical sentence types

	Mean plausibility rating (SD)		
	Nonreversible	Biased	Symmetrical
Critical sentence			
NP1-NP2 (plausible or order1)	6.23 (1.19)	6.20 (1.11)	5.63 (1.47)
NP2-NP1 (implausible or order2)	1.31 (0.88)	2.57 (1.45)	5.71 (1.38)
Ellipsis as patient			
NP1-NP3	6.18 (1.16)	5.83 (1.45)	5.56 (1.50)
NP2-NP3	1.34 (0.84)	3.01 (1.70)	5.21 (1.68)
Ellipsis as agent			
NP3-NP1	1.49 (1.19)	3.90 (1.91)	5.47 (1.51)
NP3-NP2	1.48 (1.15)	4.19 (2.00)	5.23 (1.65)

3. Relative plausibility norming study

In a separate norming study, we collected the relative plausibility ratings of each noun phrase combination (NP1-NP2, NP1-NP3, and NP2-NP3) compared directly against the role-reversed order of each combination (NP2-NP1, NP3-NP1, and NP3-NP2, respectively). The aim of this norming study is to collect the relative plausibility ratings of the two argument orders of the experimental sentences (NP1-NP2 vs. NP2-NP1) in order to include them as predictors in the main experiment, but for completeness we also collected ratings for the remaining two pairs of NP orders that contained the noun phrase in the elliptical clause (NP3). A separate set of 60 undergraduates participated in this study. We adopted the same experimental design and procedure of relative plausibility norming as Chantavarin (2021). For each item, participants read the two argument orders of each item, for example, *the dog bit the man* (NP1-NP2) and *the man bit the dog* (NP2-NP1) presented on opposite sides of a slider scale. Each item was presented with the context sentence. The left-side and right-side sentences were labelled as A and B, respectively. Participants were asked to read the context sentence and to indicate which event was more likely than the other using the slider scale. The left and right anchors were labelled “A is extremely more likely” and “B is extremely more likely”, respectively. The scale ranged from 0 (A and B are equally likely or unlikely) to 7 on the right side, and to -7 on the left side (these numeric labels were not presented to participants). The greater the absolute value of the relative plausibility rating, the greater the difference in plausibility between the pair of items. The location of each argument order was randomized so that plausible and implausible events appeared equally often on the left and right sides of the scale. Participants rated five practice trials with feedback, before rating the

126 pairs of experimental items in random order without feedback. Participants saw only one set of NP pairs for each item.

The mean relative plausibility ratings by item, NP pairs, and sentence type are presented in Figure 5. Results from this study mirrored the findings from the absolute plausibility norming study, and the mean relative plausibility ratings for each sentence type were similar to the means reported by Chantavarin (2021). As expected, for the nonreversible items participants rated the plausible order of the critical sentences (NP1-NP2) as extremely more likely than the implausible order (NP2-NP1; $M = 6.46$, $SD = 1.65$). Likewise, for the biased items, the plausible version of the experimental sentences was rated as highly more likely than the implausible version ($M = 5.53$, $SD = 2.36$). For the symmetrical items, the two argument orders of the experimental items were rated as equally plausible ($M = -0.01$, $SD = 2.69$). These ratings confirmed that each sentence type had the appropriate semantic properties by design. The remaining pairs of NP orders (i.e., the relative plausibility of NP1-NP3 compared to NP3-NP1, and NP2-NP3 compared to NP3-NP2) were not of primary interest and are therefore not discussed further.

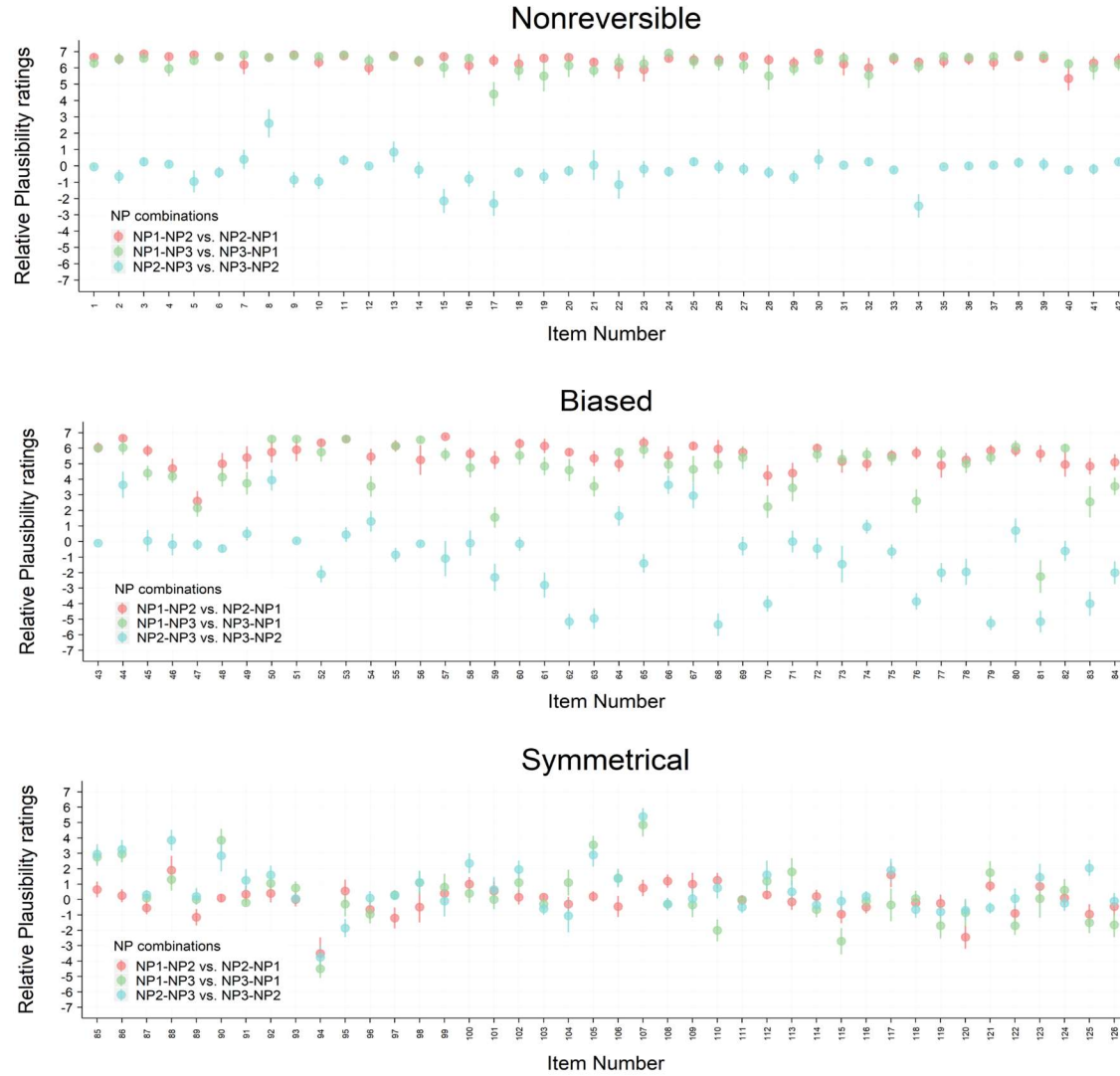


Figure 5. Mean relative plausibility ratings by items for each noun-phrase order (NP1-NP2, NP1-NP3, and NP2-NP3) compared to their role-reversed orders, for the nonreversible, biased, and symmetrical sentence types.

4. Data collection plan

The main experiment will employ a 3 (Sentence Type: nonreversible, biased, and symmetrical) x 2 (Plausibility: plausible/order1 and implausible/order2) factorial design, with 126 experimental items and approximately 100 filler items. The experimental items will be counterbalanced across six lists. The filler items will follow the same 3-sentence discourse structure as the experimental items and will contain an elliptical phrase as the third sentence.

One hundred and twenty participants will be recruited from the online SONA system at UC Davis. Participants will be randomly assigned to one of the six experimental lists. The items will be presented in a randomized order for each participant. In the experiment, participants will read each item one at a time and then respond to a comprehension question (see Figure 2 for an illustration). Each experimental item will be followed by a comprehension question about the agent of the action in the target sentence (i.e., the non-elided passive sentence), whereas the filler items will be followed by questions about the introductory context sentence (e.g., the location and time that the event took place) and about the elliptical phrase. Participants will be asked to type their response into a text box.

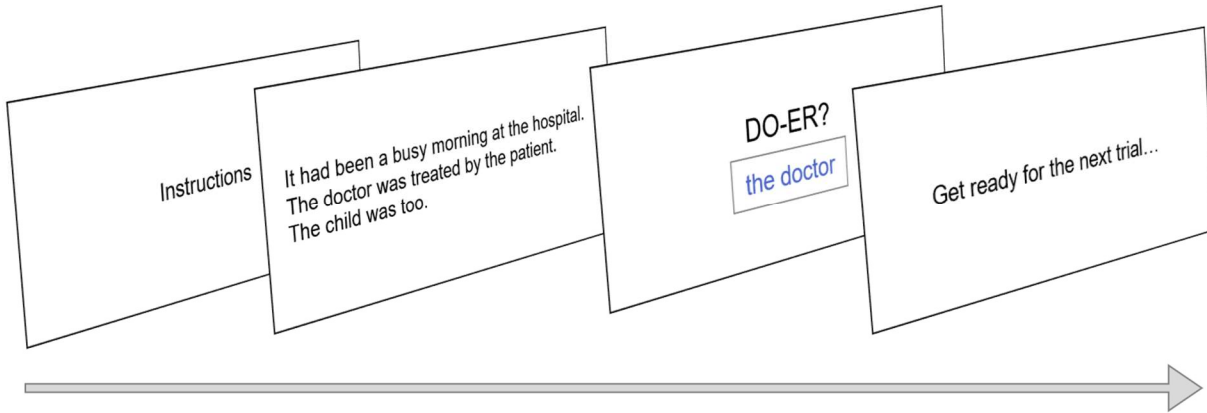


Figure 2. The planned experimental task. After going through the instructions and the practice trials, participants will read each three-sentence discourse and answer an agent identification question about the discourse they just read. Participants will be asked to type their response into a text box. Experimental items will be interspersed with filler items that have the same three-sentence discourse structure.

5. Data analysis plan

Participants who perform below 85% on the filler items will be excluded from the main analyses, and data will be collected until we obtain data from 120 participants performing above this criterion. For the analyses, we will use linear mixed-effects models that include random intercepts and slopes by items and by subjects. If a model produces a warning for singular fit or convergence error, we will iteratively remove the random effect contributing the lowest variance until the model converges with no warnings. In all analyses, each sentence type will be assessed separately.

The structure of the planned linear mixed effects models is summarized in Table 3. There will be two sets of analyses of the experimental items, addressing the two research questions as

described under Section 1.3. First, to assess whether the competing interpretations hypothesis accounts for the comprehension of passive elliptical clauses, we will test whether the target passive sentence's role-reversed plausibility and relative plausibility (obtained from the norming studies) significantly account for question answering accuracy. The two competing interpretations variables will be included as fixed effects in separate models. Second, to assess whether elided linguistic content is more prone to misinterpretations than overtly stated content, we will analyze the data obtained from this study together with the data from Chantavarin (2021) which presented the same target sentences in isolation. We will test whether comprehension accuracy (dependent variable) varies by Experiment (current study and Chantavarin, 2021) and Plausibility (plausible/order1 and implausible/order2). For completeness, we will also include the interaction between these two predictors as fixed effects (Experiment x Plausibility). Lastly, we will analyze the response times to comprehension questions as a secondary measure, following the same procedure as the planned analyses for question answering accuracy.

Table 3

Structure of linear mixed effects models for the a priori analyses

Model #	Fixed Effects	Random Effects
Research Question # 1		
1	Role-reversed plausibility	(1 Subject) + (Role-reversed plausibility Subject) + (1 Item) + (Role-reversed plausibility Item)
2	Relative plausibility	(1 Subject) + (Relative plausibility Subject) + (1 Item) + (Relative plausibility Item)
Research Question # 2		
3	Experiment*Plausibility	(1 Subject) + (Plausibility Subject) + (1 Item) + (Experiment*Plausibility Item)

Note. These three models will be run for each sentence type separately (nonreversible, biased, and symmetrical). The a priori analyses will use question answering accuracy as the dependent variable, whereas the secondary analyses will use the responses times to the comprehension questions as the dependent variable. In Model 3, the random slope of Experiment by Subjects is omitted because the set of participants differs between the current study and Chantavarin's (2021) study. Further details about these models are described in the main text.

References

- Amichetti, N. M., White, A. G., & Wingfield, A. (2016). Multiple solutions to the same problem: utilization of plausibility and syntax in sentence comprehension by older adults with impaired hearing. *Frontiers in Psychology, 7*, 789.
- Chantavarin, S. (2021). Chapter 2: Does the misinterpretation of passive sentences reflect competing interpretations? [Unpublished doctoral dissertation]. University of California, Davis.
- Christianson, K., Luke, S. G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(2), 538.
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*(4), 368-407.
- Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes, 42*(2), 205-238.
- Dabrowska, E., & Street, J. (2006). Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences, 28*(6), 604-615.
- Dalrymple, M., Shieber, S.M., Pereira, F., 1991. Ellipsis and higher-order unification. *Linguistics and Philosophy 14*, 399-452.
- Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass, 1*(1-2), 71-83.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*(2), 164-203.

- Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11-15.
- Fodor, J. D., & Inoue, A. (2000). Garden path re-analysis: Attach (anyway) and revision as last resort. In *Cross-linguistic perspectives on language processing* (pp. 21-61). Springer, Dordrecht.
- Frazier, L., & Clifton, C. (2000). On bound variable interpretations: The LF-only hypothesis. *Journal of Psycholinguistic Research*, 29(2), 125-140.
- Frazier, L., & Duff, J. (2019). Repair or accommodation? Split antecedent ellipsis and the limits of repair. *Glossa: A Journal of General Linguistics*, 4(1), 78.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6(4), 291-325.
- Garnham, A., & Oakhill, J. (1990). Mental models as contexts for interpreting texts: Implications from studies of anaphora. *Journal of Semantics*, 7(4), 379-393.
- Garnham, A. (2001). *Mental Models and the Interpretation of Anaphora (Essays in Cognitive Psychology)*. Psychology Press.
- Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, 110(20), 8051-8056.
- Hankamer, J., & Sag, I. (1976). Deep and surface anaphora. *Linguistic Inquiry*, 7(3), 391-428.
- Hardt, D. (1999). Dynamic interpretation of verb phrase ellipsis. *Linguistics and philosophy*, 185-219.
- Hoeks, J. C., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19(1), 59-73.

- Kaan, E., Wijnen, F., & Swaab, T. Y. (2004). Gapping: Electrophysiological evidence for immediate processing of “missing” verbs in sentence comprehension. *Brain and Language*, 89(3), 584-592.
- Kehler, A. (2000). Coherence and the resolution of ellipsis. *Linguistics and Philosophy*, 23(6), 533-575.
- Kertz, L. (2013). Verb phrase ellipsis: The view from information structure. *Language*, 390-428.
- 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.
- Kim, C. S., & Runner, J. T. (2018). The division of labor in explanations of verb phrase ellipsis. *Linguistics and Philosophy*, 41(1), 41-85.
- Kim, N., Brehm, L., & Yoshida, M. (2019). The online processing of noun phrase ellipsis and mechanisms of antecedent retrieval. *Language, Cognition and Neuroscience*, 34(2), 190-213.
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23-49.
- Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*, 31(5), 602-616.
- Lappin, S., 1992. The syntactic basis of ellipsis resolution. In: Berman, S., Hestvik, A. (Eds.), *Proceedings of the Stuttgart Ellipsis Workshop*.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676.
- Martin, A. E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language*, 58(3), 879-906.

- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38(3), 283-312.
- Merchant, J. (2019). Ellipsis: A survey of analytical approaches. *The Oxford Handbook of Ellipsis*, 18-46.
- Paape, D., Hemforth, B., & Vasishth, S. (2018). Processing of ellipsis with garden-path antecedents in French and German: Evidence from eye tracking. *PLoS One*, 13(6), e0198620.
- Paolazzi, C. L., Grillo, N., Alexiadou, A., & Santi, A. (2019). Passives are not hard to interpret but hard to remember: evidence from online and offline studies. *Language, Cognition and Neuroscience*, 34(8), 991-1015.
- Parker, D. (2018). A memory-based explanation of antecedent-ellipsis mismatches: New insights from computational modeling. *Glossa: A Journal of General Linguistics*, 3(1), 129.
- Phillips, C., & Parker, D. (2014). The psycholinguistics of ellipsis. *Lingua*, 151, 78-95.
- Sag, I. (1976). Deletion and logical form [Doctoral dissertation, Massachusetts Institute of Technology]. <https://dspace.mit.edu/handle/1721.1/16401>
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, 6(9), 382-386.
- Shapiro, L. P., & Hestvik, A. (1995). On-line comprehension of VP-ellipsis: Syntactic reconstruction and semantic influence. *Journal of Psycholinguistic Research*, 24(6), 517-532.
- Shapiro, L. P., Hestvik, A., Lesan, L., & Garcia, A. R. (2003). Charting the time-course of VP-ellipsis sentence comprehension: Evidence for an initial and independent structural analysis. *Journal of Memory and Language*, 49(1), 1-19.

- Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(6), 1521.
- Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In *Reading as a perceptual process* (pp. 621-648). North-Holland.
- Van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22(2), 241-255.
- Xiang, M., Grove, J., & Merchant, J. (2019). Structural priming in production through 'silence': An investigation of verb phrase ellipsis and null complement anaphora. *Glossa: A Journal of General Linguistics*, 4(1), 67.

Appendix

List of the 126 experimental items. Each item contained three sentences. The first was a filler sentence describing the context of the event (e.g., *The Friday night rush had the kitchen behind schedule*). The second was a passive sentence denoting a transitive event (e.g., *The apron_{NP2} was worn by the chef_{NP1}*). The Implausible condition was created by reversing the order of the two nouns in the passive sentence (e.g., *The chef_{NP1} was worn by the apron_{NP2}*). The last sentence contained an elliptical verb phrase, whose interpretation depended on the structure and semantic content of the preceding passive sentence (e.g., *The hat_{NP3} was ~~worn by the chef~~ too*).

Nonreversible items:

	Context	Passive-Plausible	Ellipsis
1	The Friday night rush had the kitchen behind schedule.	The apron was worn by the chef.	The hat was too.
2	Yesterday was a relaxing day on the farm.	The corn was planted by the farmer.	The wheat was too.
3	Last month, the house had a rodent problem.	The cheese was eaten by the mouse.	The cake was too.
4	The new yard was made open to the public today.	The bone was buried by the dog.	The toy was too.
5	It was a hectic afternoon in the newsroom.	The paper was reviewed by the editor.	The magazine was too.
6	The rental house had lots of issues last week.	The drain was fixed by the plumber.	The sink was too.
7	The track meet was held on a blazing summer afternoon.	The race was won by the runner.	The medal was too.
8	One spring morning, all the animals were let out to graze.	The cud was chewed by the cow.	The grass was too.
9	Everyone at the office was preparing for the meeting.	The letter was typed by the secretary.	The memo was too.
10	The studio was quiet that evening.	The picture was painted by the artist.	The mural was too.
11	The woods bustled with activity in the morning.	The wood was chewed by the termite.	The leaf was too.
12	The sun beat down on the prairie that afternoon.	The hill was built by the ant.	The tunnel was too.
13	The urgent care was packed the other day.	The X-ray was taken by the doctor.	The temperature was too.

14	The island was deserted on Monday.	The treasure was buried by the pirate.	The map was too.
15	The vaccination center was opened last Friday.	The shot was given by the nurse.	The questionnaire was too.
16	The riverbank was full of activity that morning.	The tree was gnawed by the beaver.	The log was too.
17	Work started on the new construction site today.	The dirt was pushed by the bulldozer.	The boulder was too.
18	The clothing store was busy this Wednesday.	The skirt was hemmed by the tailor.	The blouse was too.
19	The weather at the airport was cloudy that night.	The plane was flown by the pilot.	The helicopter was too.
20	Last Tuesday was a busy night at the estate.	The car was driven by the chauffeur.	The limousine was too.
21	A large audience came to watch the concert last night.	The orchestra was led by the conductor.	The choir was too.
22	The museum research room was quiet the other night.	The fossil was examined by the geologist.	The geode was too.
23	The school was preparing for the musical show.	The floor was mopped by the janitor.	The stage was too.
24	The mall reopened two weeks ago.	The dress was bought by the shopper.	The purse was too.
25	There was a commotion in the schoolyard yesterday.	The cookie was stolen by the bully.	The sandwich was too.
26	It was the last day of finals week at the university.	The exam was graded by the professor.	The assignment was too.
27	The new house was finished last year.	The room was decorated by the designer.	The backyard was too.
28	The restaurant was closed to prepare for dinner service.	The onion was chopped by the cook.	The garlic was too.
29	The royal family faced a surprise attack two years ago.	The castle was invaded by the knight.	The fortress was too.
30	The press headquarters were bombarded with work.	The article was written by the journalist.	The commentary was too.
31	The coffee shop had lots of business this morning.	The coffee was brewed by the barista.	The tea was too.
32	The architecture firm had a backlog a few days back.	The house was designed by the architect.	The backyard was too.
33	The restaurant was finally quiet after a busy night.	The table was cleaned by the busboy.	The bar was too.
34	This week the music studio was a productive place.	The song was composed by the musician.	The opera was too.
35	Last Friday the auto shop ran smoothly.	The truck was repaired by the mechanic.	The car was too.
36	The other evening the greenhouse was quite hot.	The plant was tended by the gardener.	The tree was too.
37	At dawn the training center was already open.	The barbell was lifted by the athlete.	The dumbbell was too.

38	Today the neighborhood's quiet atmosphere was disturbed.	The noise was heard by the neighbor.	The yell was too.
39	The ancient city is a popular tourist destination.	The scenery was photographed by the traveler.	The temple was too.
40	A few weeks back, numerous crimes swept the big city.	The museum was robbed by the mobster.	The bank was too.
41	Last month an uncharted wilderness was first entered.	The mountain was explored by the hiker.	The cave was too.
42	When the owner died, the old home needed a lot of work.	The roof was repaired by the contractor.	The bathroom was too.

Biased items:

	Context	Passive-Plausible	Ellipsis
1	It was a cloudy day at the dog park.	The man was bitten by the dog.	The woman was too.
2	The wind rustled in the leaves this morning.	The worm was eaten by the bird.	The grub was too.
3	After the owner left, the house was a chaotic place.	The mouse was chased by the cat.	The rat was too.
4	There was an animal attack at the village today.	The villager was protected by the soldier.	The merchant was too.
5	On Monday there was an argument at the hospital.	The doctor was sued by the lawyer.	The nurse was too.
6	The classroom was boisterous on Tuesday.	The student was quizzed by the teacher.	The friend was too.
7	The alleyway was dark that night.	The thief was pursued by the cop.	The mugger was too.
8	A new cafe opened last month.	The man was served by the waitress.	The child was too.
19	The whole household gathered in the dining room.	The cat was fed by the owner.	The dog was too.
10	Last week an investigation was opened.	The suspect was investigated by the detective.	The witness was too.
11	It had been a busy morning at the hospital.	The patient was treated by the doctor.	The child was too.
12	The mountain trail had lots of bugs in the morning.	The mosquito was killed by the hiker.	The tick was too.
13	It was a chaotic afternoon at the stable.	The rider was thrown by the horse.	The trainer was too.
14	The woods were foggy that evening.	The deer was shot by the hunter.	The bear was too.
15	The pond was peaceful that afternoon.	The fly was eaten by the frog.	The worm was too.
16	The haunted mansion is an exciting place to visit.	The boy was scared by the ghost.	The mother was too.
17	It was raining on the ocean that day.	The fish was caught by the angler.	The eel was too.
18	The busy street was a crime scene the other day.	The citizen was arrested by the officer.	The bystander was too.

19	The other day's earthquake collapsed the bridge.	The child was carried by the paramedic.	The sibling was too.
20	A meeting was held at the office on Friday.	The student was advised by the counselor.	The parent was too.
21	Yesterday the family walked home together.	The toddler was scolded by the parent.	The teenager was too.
22	The prairie is an unforgiving environment.	The rabbit was devoured by the snake.	The mole was too.
23	The district court was in session yesterday.	The terrorist was sentenced by the judge.	The arsonist was too.
24	Last Sunday the weather at the park was lovely.	The child was watched by the babysitter.	The baby was too.
25	The university reopened for the new semester.	The class was instructed by the lecturer.	The discussion was too.
26	The firm was productive on Wednesday.	The intern was hired by the supervisor.	The secretary was too.
27	There was a threat called in at the town hall.	The mayor was guarded by the cop.	The senator was too.
28	The studio had a shoot that night for the new film.	The model was discovered by the scout.	The photographer was too.
29	Police arrived on the crime scene at midnight.	The perpetrator was accused by the victim.	The accomplice was too.
30	The amphitheater was packed that evening.	The musician was applauded by the audience.	The composer was too.
31	The national park is quite confusing to navigate.	The tourist was guided by the ranger.	The camper was too.
32	That morning the beach was a dangerous place.	The surfer was saved by the lifeguard.	The swimmer was too.
33	Two days ago a fire burned down the apartment building.	The family was rescued by the firefighter.	The neighbor was too.
34	The family enjoyed a relaxing afternoon at home.	The baby was tickled by the mother.	The child was too.
35	The capitol building was a hive of activity that day.	The president was filmed by the cameraman.	The ambassador was too.
36	There was a show at the theater last night.	The crowd was entertained by the magician.	The reviewer was too.
37	The school office was occupied today.	The pupil was reprimanded by the principal.	The teacher was too.
38	The competition show was a very close call.	The recipient was awarded by the host.	The runner-up was too.
39	There was an altercation at the work site yesterday.	The apprentice was fired by the carpenter.	The contractor was too.
40	Last month's diplomatic meeting went well.	The prince was escorted by the bodyguard.	The duke was too.
41	A special guest arrived at the airport with his relatives.	The visitor was welcomed by the local.	The family was too.

42	It was a rainy morning in the war zone.	The troop was led by the commander.	The pilot was too.
----	---	-------------------------------------	--------------------

Symmetrical items:

	Context	Passive-Order1	Ellipsis
1	A fight broke out this morning in the classroom.	The girl was kicked by the boy.	The teacher was too.
2	The family reunited at the airport on Sunday.	The boyfriend was kissed by the girlfriend.	The baby was too.
3	The birthday party yesterday was a joyful affair.	The brother was hugged by the sister.	The uncle was too.
4	The annual conference took place last Wednesday.	The chairman was introduced by the committee.	The sponsor was too.
5	A family played at the nearby playground today.	The girl was called by the woman.	The boy was too.
6	Child protection services were visiting several homes.	The woman was visited by the man.	The child was too.
7	The awards show yesterday was well attended.	The director was recognized by the producer.	The screenwriter was too.
8	The cultural fair last weekend encouraged communication.	The rabbi was approached by the priest.	The imam was too.
9	Gossip about the shop owners circulated in the small town.	The baker was despised by the butcher.	The florist was too.
10	The Major League draft was held two weeks ago.	The player was chosen by the team.	The coach was too.
11	The store was filled with people shopping for the holidays.	The customer was thanked by the clerk.	The manager was too.
12	The school office was full of people that morning.	The parent was greeted by the teacher.	The principal was too.
13	The family home was especially cozy on Christmas.	The son was adored by the mother.	The daughter was too.
14	The apartment got a lot of attention after the open house.	The buyer was faxed by the realtor.	The owner was too.
15	The sale of the antique was cancelled before noon.	The client was phoned by the broker.	The seller was too.
16	Political differences led to an argument at the party.	The host was insulted by the guest.	The caterer was too.
17	The scouting agency set up an interview this Thursday.	The photographer was met by the model.	The agent was too.
18	A parade was held in the capital on Easter.	The pope was embraced by the prime minister.	The cardinal was too.
19	Today, the magic academy celebrated the graduating class.	The wizard was praised by the witch.	The sorcerer was too.

20	Thursday's hiring fair raised some legal questions.	The company was contacted by the applicant.	The lawyer was too.
21	The funeral home was a somber place today.	The mother was consoled by the father.	The child was too.
22	A shootout had broken out in the downtown area.	The sheriff was killed by the gangster.	The arsonist was too.
23	The company headquarters was busy on Wednesday.	The salesman was helped by the accountant.	The client was too.
24	The classroom got rowdy when the teacher left.	The girl was teased by the boy.	The friend was too.
25	There was a problem with the manuscript this morning.	The author was consulted by the editor.	The publisher was too.
26	There will be a formal dinner at the hotel next Tuesday.	The diplomat was invited by the governor.	The mayor was too.
27	Career Day at the school led to new friendships.	The astronaut was respected by the policeman.	The firefighter was too.
28	Online school made the children restless this past year.	The sister was pestered by the brother.	The mother was too.
29	The castle was invaded in the middle of the night.	The intruder was defeated by the guard.	The duke was too.
30	The soccer match last night was intense.	The opponent was tackled by the player.	The teammate was too.
31	The construction site was working overtime the other day.	The worker was assisted by the foreman.	The engineer was too.
32	The Halloween party on Friday got a bit out of hand.	The vampire was kicked by the zombie.	The werewolf was too.
33	The mall was an active crime scene on Sunday.	The shooter was confronted by the trooper.	The employee was too.
34	A few years back, rumors of infidelity spread around the town.	The wife was accused by the husband.	The neighbor was too.
35	There was a scandal at the military base last month.	The spy was shot by the inspector.	The general was too.
36	An altercation occurred at Monday's art sale.	The reviewer was mocked by the artist.	The buyer was too.
37	The other day was an eventful one at the zoo.	The leopard was attacked by the tiger.	The crocodile was too.
38	There was an altercation at the bar yesterday.	The magician was deceived by the prankster.	The swindler was too.
39	The investment bank's payday was on Thursday.	The cashier was paid by the banker.	The trader was too.
40	Last night the movie won a prestigious award at the gala.	The actor was congratulated by the actress.	The director was too.
41	After a long day, the restaurant staff sat down to eat.	The bartender was served by the waiter.	The hostess was too.
42	It was a wild night in the jungle.	The lion was hunted by the bear.	The cheetah was too.

CHAPTER 4

Conclusions

The goal of this dissertation was to examine the competing interpretations hypothesis as a possible mechanism underlying the misinterpretation of noncanonical structures. If the language comprehension system computes more than one analysis at once, as assumed under parallel-competitive models of parsing (Hagoort, Baggio, & Willems, 2009; Kos, Vosse, Van Den Brink, & Hagoort, 2010; also see McRae, Spivey & Tanenhaus, 1998; MacDonald, Pearlmutter, & Seidenberg, 1994), we might expect a competition between the veridical interpretation and other alternative interpretations that are highly plausible, which may include interpretations that are not licensed by the syntactic structure. The semantic cues from the verb and its arguments may activate the role-reversed interpretation, and this alternative, non-veridical interpretation may occasionally be selected when comprehenders rely on semantic-heuristic strategies when processing complex structures such as passive sentences and ellipses. Unlike previous studies that examined the likelihood of misinterpretations as a function of the veridical interpretation's plausibility, we assessed how the plausibility of an alternative, role-reversed interpretation influences the final interpretation. This innovation allowed a more direct investigation of the potential mechanisms by which a non-veridical interpretation is selected instead of the linguistically correct one.

The experiments described in Chapter 2 tested the competing interpretations hypothesis in the comprehension of passive sentences. Results showed that misinterpretation errors were greater as the plausibility of the role-reversed interpretation increased, but this was found only for the sentences that denoted a syntax-semantics conflict, and not for the sentences that contained two equally compelling meanings. These findings provide evidence against the competing

interpretations hypothesis that assumed parallel activation and competition between multiple parses, and instead supports models of parsing that assume a commitment to a single analysis that may later be revised when a semantic anomaly is encountered (Frazier & Fodor, 1978; Fodor & Inoue, 1998; also see Van Gompel, Pickering, & Traxler, 2000, 2001). We favor this interpretation of the results because it is the most parsimonious account of the findings across the three sentence types that were tested (i.e., nonreversible, biased, and symmetrical sentences).

This dissertation is a preliminary investigation of the competing interpretations hypothesis, and the reported experiments were concerned with the factors that influenced comprehenders' final interpretation of noncanonical sentences. Investigating the content of the representations that are built during comprehension are an important complement to the data collected from online measures such as eye tracking and event-related potentials (Ferreira & Yang, 2019). Although our findings indicated that the competing interpretation's plausibility does not influence comprehension accuracy on the symmetrically plausible sentences, an open question for future research is to investigate whether and how competing interpretations impact the real-time processing dynamics of noncanonical structures. More generally, this dissertation also revisits the role of integration processes and event inferencing in language comprehension as featured in the older theoretical approaches (e.g., Kintsch, 1988; Gernsbacher, 1991; Singer, Graesser, & Trabasso, 1994; McKoon & Ratcliff, 1998), which are critical to a comprehensive account of language comprehension. In contrast to current approaches to parsing that emphasize anticipatory processing (e.g., Hale, 2001; Levy, 2008; Gibson, 2000; Van Petten & Luka, 2012; Kuperberg & Jaeger, 2016), these older traditions emphasized the integration of the current linguistic input with background knowledge in long-term memory (e.g., event-based knowledge), which supports a

coherent semantic representation of the text or conversation (for further discussion, see Ferreira & Chantavarin, 2018). The competing interpretations hypothesis proposed in this dissertation opens up avenues for future research regarding the nuanced role of semantic-heuristic strategies in the completeness of the linguistic representations that are generated during the comprehension of noncanonical sentences.

In closing, the experiments reported in this dissertation investigated how everyday language input is interpreted, including semantically anomalous sentences which can result from imperfections in language production and perception. Specifically, the studies probed the accuracy of people's memory of syntactically complex sentences, in order to characterize the factors that influence this comprehension process. Given that communicating about events is ubiquitous in everyday language use, it is important to investigate the cognitive processes that enable our understanding of who did what to whom from the language input we receive. Beyond sentence processing in healthy adults, the studies reported here have implications in cognitive aging and disease, such as the extent to which older adults and clinical populations rely on prior knowledge to process linguistic events. The current work can also inform our understanding of language acquisition in children and in second language learners, providing insight into the processes that support how noncanonical constructions such as passives and ellipses are learned and remembered, and how prior knowledge affects these processes.

References

- Ferreira, F., & Chantavarin, S. (2018). Integration and prediction in language processing: A synthesis of old and new. *Current Directions in Psychological Science*, 27(6), 443-448.
- Ferreira, F., & Yang, Z. (2019). The problem of comprehension in psycholinguistics. *Discourse Processes*, 56(7), 485-495.
- Fodor, J. D., & Inoue, A. (1998). Attach anyway. In *Reanalysis in sentence processing* (pp. 101-141). Springer, Dordrecht.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6(4), 291-325.
- Gernsbacher, M. A. (1991). Cognitive processes and mechanisms in language comprehension: The structure building framework. *Psychology of Learning and Motivation*, 27, 217-263.
- Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium* (pp. 95-126). Cambridge, MA: MIT Press.
- Hagoort, P., Baggio, G., and Willems, R. (2009). "Semantic unification," in *The New Cognitive Neuroscience*, ed. M. S. Gazzaniga (Cambridge, MA: MIT Press), 56, 1-18.
- Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL* (Vol. 2, pp. 159-166).
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163.
- Kos, M., Vosse, T. G., Van Den Brink, D., & Hagoort, P. (2010). About edible restaurants: conflicts between syntax and semantics as revealed by ERPs. *Frontiers in Psychology*, 1, 222.

- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. *Language, Cognition and Neuroscience*, 31(1), 32-59.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological review*, 101(4), 676.
- McKoon, G., & Ratcliff, R. (1998). Memory-based language processing: Psycholinguistic research in the 1990s. *Annual Review of Psychology*, 49(1), 25-42.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38(3), 283-312.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371.
- Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In *Reading as a perceptual process* (pp. 621-648). North-Holland.
- Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, 45(2), 225-258.
- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2), 176-190.