UNIVERSITY OF CALIFORNIA SAN DIEGO

Towards a Referential Theory of Ellipsis

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

in

Linguistics

by

Till Poppels

Committee in charge:

Professor Andrew Kehler, Chair
Professor Roger Levy, Co-Chair
Professor David Barner
Professor Ivano Caponigro
Professor Victor Ferreira

2020

The dissertation of Till Poppels is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California San Diego

2020

DEDICATION

To my mother, Regina Poppels,

for teaching a little boy the joy of discovery.

EPIGRAPH

Your most important work is always ahead of you, never behind you.

—Stephen R. Covey

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am grateful to a great number of people who have helped and supported me along the way—without you, none of this would have been possible.

First, I have to thank my advisors Andy Kehler and Roger Levy who, more than anyone else, have shaped the way I think about language and science. Both of them have always been incredibly generous with their time and witnessing their enthusiasm about the scientific study of language has been a source inspiration to me throughout my time in grad school. Andy's persistent encouragement to never lose sight of the big picture and to keep pursuing the most exciting questions about language has taught me to resist the temptation to "code up" superficial explanations and to aspire to understand phenomena in deep, fundamental ways. He has also taught me the importance of telling a story when I present my research, which is no doubt one of the most generally applicable lessons I have learned in grad school.

Roger has taught me to think about language as a window onto some of the most vexing and intriguing puzzles in the study of the human mind. From the very first days of my grad school experience, his welcoming and intellectually rigorous "no BS" lab culture was a source of inspiration to me. It showed me how fascinating psycholinguistic and computational language research could be, and seeing how much there was for me to learn was deeply motivating. Roger also invited me to spend a year at the Brain and Cognitive Sciences department at MIT, which was a thoroughly positive and formative experience for me.

Thanks are also due to my other committee members: Dave Barner, Ivano Caponigro, and Vic Ferreira. I have had so many interesting and inspiring discussions at the *Semantics Babble* meetings that Dave and Ivano organize, and both of them have provided invaluable feedback on my research at various stages. Their advice was always incredibly insightful and has shaped the way I think and talk about my work. Vic has also been a great source of support throughout my time at UCSD, both as a researcher and over beers and dinners at CUNY and beyond.

I am also grateful to many other people who have made my time in grad school a fun and

engaging experience: Gaby Caballero for introducing me to the wonders of field work; Grant Goodall for engaging and emotionally soothing lab meetings; Leon Bergen for the most enjoyable and interesting TA-ing experiences; Savi Namboodiripad for being the greatest conference buddy anyone could hope for; Helena Aparicio for being one of the kindest and most inspiring people I know; Titus von der Malsburg for being an incredible colleague and co-author; Alycia Randoll, Lucie Wiseman, Rachel Pekras, and Silas Horton for their prompt and always friendly help with any administrative and technical issue I ran into at UCSD; Anne Abeillé and Philip Miller for inviting me to Paris and making the transition out of grad school a breeze. I would also like to acknowledge the Majority Report community—especially Sam, Michael, Jamie, Matt, and Brandon—whose insightful and delightfully comedic political commentary has kept me afloat during these most unsettling times in the United States.

Finally, I would like to thank my family—Olga, Regina, Nils, Anne, Hagen, Marita, Hildegard, Rolf, Mizzi, and Peter—and my best friends of 20 years, Mirko and Tobias. As long as I can remember, you have always encouraged me to pursue whatever seemed most interesting at the time, and I would not have ended up where I am today if it hadn't been for your unconditional acceptance and unwavering emotional support. I love and appreciate you.

Several chapters in this thesis reflect collaborative work with Andy Kehler. The material in Chapter 3 has been published in: Poppels, T., & Kehler, A. (2019). Reconsidering asymmetries in voice-mismatched verb phrase ellipsis. *Glossa: A Journal of General Linguistics*, *4*, 1–22. doi:http://doi.org/10.5334/gjgl.738. Chapter 4 has been published in Poppels, T., & Kehler, A. (2018). Overcoming the identity crisis: Novel evidence for a referential theory of verb phrase ellipsis. In *Proceedings of the Annual Meeting of the Chicago Linguistics Society* (Vol. 53, pp. 403–4017). Chapters 5 and 6 are presently being prepared for submission in collaboration with Andy Kehler. In each case, the thesis author is the primary author.

VITA

2014	MA in Mind & Language University of Edinburgh, UK
2014-2017	Graduate Student Researcher, Computational Psycholinguistics Lab University of California, San Diego, USA
2017-2018	Visiting Student Researcher, Computational Psycholinguistics Lab Massachusetts Institute of Technology, USA
2020	Ph.D. in Linguistics University of California, San Diego, USA

PUBLICATIONS

Titus von der Malsburg, Till Poppels, and Roger Levy (2020). "Implicit Gender Bias in Linguistic Descriptions for Expected Events: The Cases of the 2016 US and 2017 UK Elections." *Psychological Science*.

Till Poppels and Andrew Kehler (2019). "Reconsidering Asymmetries in Voice-Mismatched Verb Phrase Ellipsis." *Glossa: A Journal of General Linguistics*.

Till Poppels and Andrew Kehler (2018). "Overcoming the Identity Crisis: Novel Evidence for a Referential Theory of Verb Phrase Ellipsis." In *Proceedings of the 53rd Annual Meeting of the Chicago Linguistics Society*.

Till Poppels and Roger Levy (2016). "Structure-Sensitive Noise Inference: Comprehenders Expect Exchange Errors." In *Proceedings of the 38th Annual Conference of the Cognitive Science Society.*

Till Poppels and Roger Levy (2015). "Resolving Quantity and Informativeness Implicature in Indefinite Reference." In *Proceedings of the 20th Amsterdam Colloquium*.

ABSTRACT OF THE DISSERTATION

Towards a Referential Theory of Ellipsis

by

Till Poppels

Doctor of Philosophy in Linguistics

University of California San Diego, 2020

Professor Andrew Kehler, Chair
Professor Roger Levy, Co-Chair

Ellipsis is a pervasive phenomenon across the world's languages, and it is easy to see why: it allows speakers to omit certain parts of their utterances while nonetheless conveying their full meaning, which contributes to making linguistic communication highly efficient. There is broad consensus that elliptical utterances depend on the context in some way, but the nature of this dependency remains controversial. Broadly speaking, theories of ellipsis fall into one of two camps that make fundamentally distinct architectural assumptions about the mechanisms that enable ellipsis: IDENTITY theories, which posit that material can be elided only if it is identical to a linguistic antecedent; and referential theories, which assume that ellipsis is enabled by the same

underlying mechanism that governs other forms of discourse reference. This thesis draws on evidence from a series of experiments that investigate two prominent types of ellipsis: VP-ellipsis and sluicing. It focuses on cases of mismatch between the elided material and its antecedent in which the meaning of the ellipsis clause is not reducible to the linguistic context and must instead be inferred. Taken together, this investigation makes three main contributions. First, it raises novel challenges for IDENTITY theories by demonstrating that ellipsis can be felicitous even in the face of extreme mismatch. Secondly, it provides new evidence in support of referential theories by showcasing the possibility of inferential ellipsis resolution—a hallmark property of discourse reference. Finally, and most importantly, it contributes new adequacy criteria for linguistic theories aimed at explaining the nature of the linguistic and non-linguistic context and how it interfaces with context-dependent linguistic devices.

Chapter 1

Synopsis

The fact that languages around the world provide ways for speakers to elide certain parts of their utterances contributes to making linguistic communication highly efficient, as illustrated in (1): instead of repeating material that is redundant in the context of the surrounding discourse (shown in brackets), it can be omitted without changing the meaning of the utterance. In order for communication to succeed, comprehenders must, of course, be able to fill in the missing pieces, which they usually do effortlessly:

(1) a. Someone solved the problem, but I don't know who (solved the problem).

b. Did you not tell your friends about the game today? —I did, but I forgot to tell them where (the game is).

Given the communicative benefits of ellipsis, one might expect it to be a simple affair: whenever some part of an utterance is redundant in some way, eliding it should be permissible if not preferred. As it turns out, however, defining the notion of redundancy that ellipsis requires is so complicated that more than five decades of linguistic research have been insufficient to capture, let alone explain, the conditions that make ellipsis felicitous. Consider the following examples, which differ only slightly from the ones in (1) but are nonetheless much less acceptable.

(2) a. The problem was solved, but I don't know who #(solved the problem).

b. Did you not tell your friends about the game today? —I did, but I forgot to tell them how long #(the game is).

As I argue at length in Chapter 2, I believe that one of the reasons that ellipsis remains elusive to date is that the literature has been experiencing an "IDENTITY crisis"—an overreliance on the axiomatic assumption that material can be elided only if the linguistic context provides an identical copy of it. This assumption, which is at the heart of an influential line of theories known as IDENTITY theories, provides a straightforward explanation for the acceptability of some uses of ellipsis and the *un*acceptability of others. For example, the context in (1a) is expected to enable ellipsis because it provides an identical copy of the to-be-elided material, and the unacceptability of (2a) similarly follows from the fact that the antecedent has been passivized and therefore no longer satisfies the IDENTITY requirement.

But there are also many counterexamples that suggest that the IDENTITY condition is neither necessary nor sufficient for characterizing the distribution of ellipsis. Consider (1b), which appears to permit the elision of the clause *the game is* even though the context only provides an antecedent NP—*game*—that is neither syntactically nor semantically identical to the elided material. Even more puzzling is the fact that the exact same antecedent-ellipsis pair appears to be infelicitous in (2b), even though it only differs from (1b) with respect to the wh-phrase *how long*. If, as is widely assumed in the literature, the acceptability of ellipsis is reducible to some relation between the elided material and its linguistic antecedent, the contrast between (1c) and (2c) remains unexplained.

This thesis reexamines some of the fundamental assumptions in the ellipsis literature, focusing in particular on VP-ellipsis and sluicing in English. It makes the case that in order to overcome the IDENTITY crisis, we should turn to an alternative tradition of theorizing that construes ellipsis as a form of discourse reference. According to referential theories, elliptical ut-

terances are governed by the same underlying mechanism that supports other discourse-referential devices, such as pronouns. Chapter 2 begins by reviewing the history of both IDENTITY theories and referential theories, outlines arguments for and against both traditions, and sketches open questions to be addressed in future research. Against this backdrop, the subsequent chapters focus on various types of mismatches between the elided material and its linguistic antecedent, which speak to both types of theories: for IDENTITY theories, they represent new challenges that constrain viable definitions of IDENTITY; and for referential theories, they provide important adequacy criteria for modeling the inferences that enable reference resolution.

Chapter 3 kicks off the investigation by presenting a series of experiments that investigate sources of gradience associated with voice-mismatched VP-ellipsis and argues that the results are inconsistent with IDENTITY-based explanations. Chapter 4 then goes on to explore cases of VP-ellipsis that involve knowledge-driven inferences, which are exemplified in (3). In those cases, the ellipsis clause acquires a meaning that is related to the event denoted by the antecedent VP, but not identical to it. I will argue that these examples are also beyond the explanatory reach of IDENTITY theories and that the inferences they involve resemble those associated with non-elliptical forms of discourse reference.

(3) a. A: Can I borrow your textbook over the weekend?
B: I can't (lend it to you), I'll need it myself.

b. Spectator: Can I see that card trick one more time?
Magician: I can't (show it to you), sorry.

Chapter 5 then pivots to sluicing and discusses evidence for acceptable argument-structure mismatches based on tough movement and the passive/active voice alternation, as in (4). The results from these experiments are inconsistent with some, but not all, existing formulations of IDENTITY, and they raise novel questions for referential theories of sluicing.

(4) a. Brake fluid is easy to replace if you know how (to replace it).

b. The problem hasn't been solved because no one knows how (to solve it).

Finally, Chapter 6 examines even more disruptive examples of sluicing that involve mismatches of various sorts, exemplified in (5) and (6).

(5) Fan: Can I get a few autographs?

Manager: Sure, how many (do you want/need)?

(6) Regarding Trump's impeachment, the only question was...

a. ...when (he would be impeached).

b. ...why #(he would be impeached).

c. ...who ##(he would be impeached by).

Some of these examples will turn out to be highly acceptable, while others exhibit a considerable amount of gradience, raising new challenges for IDENTITY theorists and referential theorists alike.

Chapter 2

Background

Ellipsis is a cross-linguistically prevalent phenomenon and it is easy to see why: it allows speakers to omit information that is in some way available in the context, making linguistic utterances much more efficient than they would be without the possibility of ellipsis. Since languages are optimized for efficient communication in many ways (Gibson et al., 2019), it is perhaps unsurprising that there are many different types of ellipsis. Some of the ways to elide material in English are exemplified in (7).

(7) a. **Sluicing**: I remember the paper was about ellipsis, but I don't remember which type of ellipsis (the paper was about).

b. **VP-ellipsis**: Jackie thought she might win the jackpot, but sadly she didn't (win it).

c. **NP-ellipsis**: I don't know about Bill's (internet connection), but my internet connection is pretty slow.

d. **Null Complement Anaphora**: Millions of people were going to lose their health insurance, but the Republicans didn't care (that millions were going to lose their health insurance).

e. **Fragment answer**: Who do you think will clinch the nomination? —Bernie Sanders (I think will clinch the nomination).

f. **Gapping**: Julie called Mike and Sarah (called) Ben.

g. **Pseudogapping**: She talked about her own research more than she did (talk about) anyone else's.

In all of these examples, the bracketed material is optional, i.e. it can be felicitously omitted without changing the meaning of the utterance. For want of a pre-theoretical definition of "ellipsis," I take it to be a set of grammatical devices that enable the sort of "meaning-preserving optionality" illustrated in (7):[1]

(8) **Ellipsis:** the omission of otherwise grammatically mandated material that (a) preserves the meaning of the utterance, and (b) results in an acceptable utterance in its own right.

This thesis focuses on two types of ellipsis: a type of clausal ellipsis, known as "sluicing," exemplified in (7a) and further illustrated in (9); and VP-ellipsis, shown in (7b) and (10).

(9) a. When Jimmy decided to go, he told no one where (he would go).

b. A: I can't come to your party.
B: Why (can you not come to my party)?

c. Customer: Can I have one of those donuts over there?
Counterperson: Sure. Which one (would you like)?

(10) a. The defender didn't expect the shot, but the goalkeeper did (expect the shot).

b. Nurses nation-wide deserve a pay raise, and cleaning personnel do (deserve a pay raise), too.

c. Spencer was at home, and Dillon was (at home), too.

[1] Classifying all of these cases as "ellipsis" is by no means uncontroversial (but see Ginzburg & Miller, 2019, for a similarly broad definition that covers all of these examples and more). In fact, in the literature the term "ellipsis" is often used in a theory-laden way that presupposes, for example, that elliptical utterances are derived from their non-elliptical counterparts or that the elided material is fully represented at the level of syntax and only "missing" phonologically. By contrast, I use the term "ellipsis" as a purely descriptive label that is not intended to imply a certain analysis or even that all types of ellipsis are governed by a single mechanism.

Sluicing is a cross-linguistically prevalent form of ellipsis that involves the omission of clauses that are embedded under interrogative wh-phrases. VP-ellipsis, on the other hand, involves the omission of sub-clausal material, which, despite what its name suggests, is not limited to verb phrases: it applies to any material following an auxiliary.[2]

One aspect that all forms of ellipsis have in common is that the meaning of the elided material must be provided by the context in some way. Indeed, in the absence of such contextual support elliptical utterances are neither fully interpretable nor felicitous:

(11)
a. Bill didn't #(...). — VP-ellipsis
b. Sarah #(...). — Fragment answer
c. I don't know #(...). — Null Complement Anaphora
d. No one knows why #(...). — Sluicing
e. Devin's #(...) didn't show up. — NP-ellipsis

While it is uncontroversial *that* the use of ellipsis is context-dependent, theories of ellipsis differ in terms of their answer to a fundamental theoretical question that goes back at least to Hankamer and Sag (1976): What is the nature of the dependency between an elliptical utterance and its context, and which part of the language architecture is responsible for resolving it?

According to the seminal work of Hankamer and Sag (1976) and Sag and Hankamer (1984), which I will jointly refer to as "H&S," there are two natural classes of anaphoric expressions, which are "interpreted by entirely different means" (Sag & Hankamer, 1984, p. 338): model-interpretive ("deep") anaphora, which involves reference to a representation in the interlocutors' shared mental model of the discourse (known as the "discourse model"); and "surface" anaphora, which depends directly on a linguistic constituent known as the "antecedent," without being mediated by the discourse model. H&S argue that these different classes of anaphoric

[2] In light of this fact, Miller and Pullum (2013) propose the descriptively more accurate label "post-auxiliary ellipsis." For consistency with the majority of the literature, however, I will continue to use the term "VP-ellipsis."

devices have distinctive diagnostic properties. First, model-interpretive anaphoric expressions can refer exophorically, i.e. in the absence of a linguistic antecedent, while surface anaphors, which depend directly on the linguistic antecedent, cannot. Based on this diagnostic, Hankamer and Sag (1976) draw on data like (12a-b) to argue that certain elliptical expressions, including VP-ellipsis and sluicing, must be forms of surface anaphora, whereas other anaphoric expressions, such as *do it* in (12c), involve model-interpretive anaphora.

(12) [Context: Hankamer produces a gun, points it offstage and fires, whereupon a scream is heard]

a. Sag: Jesus, I wonder who #(has been shot). (Hankamer & Sag, 1976, ex. 43)

b. Sag: Jorge, you shouldn't have #(fired the gun). (Sag & Hankamer, 1984, ex. 5d)

c. Sag: Jorge, you shouldn't have done it. (Sag & Hankamer, 1984, ex. 5e)

Secondly, H&S suggest that surface anaphors, but not model-interpretive referring expressions, are sensitive to morphosyntactic properties of their linguistic antecedents. That diagnostic also supports their conclusion that VP-ellipsis and sluicing are instances of surface anaphora, since they show sensitivity to voice mismatches between the ellipsis clause and the antecedent clause, whereas *do it*, once again, behaves differently:

(13) Maggie was made fun of, but she couldn't see...

a. ...who #(made fun of her). Sluicing

b. ...who did #(make fun of her). VP-ellipsis

c. ...who did it. *do it* anaphora

H&S's bipartite theory of anaphoric context dependence, which construes ellipsis as fundamentally different from other (model-interpretive) anaphoric devices, set the stage for a long-standing and influential tradition of so-called "IDENTITY theories" of ellipsis (Sag, 1976;

Fiengo & May, 1994; Chung, Ladusaw, & McCloskey, 1995; Merchant, 2001; Chung, 2006; Chung, Ladusaw, & McCloskey, 2011; Chung, 2013; Rudin, 2019, among many others). According to these theories, elliptical constructions like VP-ellipsis and sluicing are governed by an ellipsis-specific mechanism that requires the elided material to be identical to its antecedent. By contrast, running counter to H&S's conclusion, referential approaches to ellipsis have pursued the hypothesis that ellipsis is governed by the same mechanisms of discourse reference that support other anaphoric expressions (Wasow, 1972; Webber, 1978; Hardt, 1993; Kehler, 2000; Jäger, 2005; Barker, 2013, and many others).

The remainder of this chapter will discuss both of these theoretical approaches in detail and consider arguments for and against each. Section 2.1 surveys different flavors of IDENTITY theories that have been proposed over time and explores the theoretical and empirical implications of this approach. Section 2.2 will then discuss referential theories of ellipsis by extending H&S's original argumentation to other diagnostic properties beyond exophora and mismatch facts. I will argue (i) that both VP-ellipsis and sluicing do exhibit all of the hallmark features of discourse reference, and (ii) that non-elliptical referring expressions do show the kind of sensitivity to the morphosyntactic form of their antecedents that H&S considered unique to ellipsis. Finally, Section 2.3 offers a comparison between IDENTITY theories and referential theories of ellipsis, summarizes the arguments on both sides, and situates the contributions of each chapter in this thesis within this theoretical landscape.

2.1 IDENTITY theories of ellipsis

Starting from the consensus position that elliptical utterances depend on the context in *some* way, the central claim behind the class of theories that I refer to as "IDENTITY theories" can be understood as reducing that dependency to a particular part of the context:

(14) **Central claim:** the context-sensitivity of elliptical utterances is reducible to a yet-to-be-

defined "IDENTITY" relation between the ellipsis site and its linguistic antecedent.

Some version of this assumption has served has the starting point for a large number of theoretical proposals (Sag, 1976; Dalrymple, Shieber, & Pereira, 1991; Chung et al., 1995; Merchant, 2001; Elbourne, 2008; Chung, 2013; Van Craenenbroeck & Merchant, 2013; Lipták, 2015; Rudin, 2019; Vicente, 2019, among many others). In fact, it is often presupposed axiomatically and merely frames the main research question that concerns IDENTITY theories of ellipsis, which Lipták (2015) refers to as the "quest for identity": assuming that *some* identity relation must hold between the ellipsis site and its antecedent, how does this relation have to be defined in order to correctly classify elliptical utterances as acceptable or unacceptable. As outlined below, this research program is complicated by the recalcitrant nature of ellipsis data, which has spurred a cottage industry of research that follows what I call an "incremental fine-tuning approach": the definition of the grammatically mandated IDENTITY condition is incrementally revised to accommodate increasingly large sets of previously problematic data. Section 2.1.1 illustrates this approach by introducing prototypical examples of IDENTITY theories that have been proposed over time, along with the empirical considerations that motivated them. Sections 2.1.2 and 2.1.3 will then briefly describe two other "fine-tuning" strategies that pursue the same goal but exploit theory-external degrees of freedom, rather than targeting the definition of IDENTITY itself.

2.1.1 The IDENTITY crisis: "fine-tuning" the definition of IDENTITY

At first glance, the notion that ellipsis requires identity between the elided material and its antecedent is straightforward enough. Consider the following minimal pairs:

(15) a. A: Nina talked to someone.
B: Oh yeah? I wonder...
(i) ...who (Nina talked to)?

(ii) ...who #(Nina had an argument with)?

b. A: Nina had an argument with someone.

B: I'm surprised. Are you sure...

(i) ...she did #(talk to someone)?

(ii) ...she did (have an argument with someone)?

To explain the impossibility of (15a-ii) and (15b-i) in terms of IDENTITY, we start with the assumption that the meaning of the ellipsis clause is derived (in the usual way) from the elided material, and I will use ~~strike-out~~ notation to make that assumption explicit:

(16) a. A: Nina talked to someone.

B: Oh yeah? I wonder...

(i) ...who ~~she talked to~~?

(ii) #...who ~~she had an argument with~~?

b. A: Nina had an argument with someone.

B: I'm surprised. Are you sure...

(i) #...she did ~~talk to someone~~?

(ii) ...she did ~~have an argument with someone~~?

Since IDENTITY theories permit ellipsis only when the to-be-elided material is identical to its antecedent, the interpretations in (16a-ii) and (16b-i) can only be conveyed non-elliptically (and, in fact, there is nothing wrong with those interpretations in the absence of ellipsis). Conversely, when there is an exact match between the two, ellipsis is licensed and the corresponding interpretation becomes available. However, as we will see over the course of this section, requiring an exact match at all levels of representation is overly strict since various types of "mismatches" do not, in fact, prevent ellipsis.

Consider the following examples, which appear to license ellipsis in spite of a lexical

mismatch between the elided material and its antecedent (underlined):

(17) a. They arrested Alex, though he thought they wouldn't ~~arrest him~~.

(Merchant, 2001, ex. 38)

b. A: I will call my sister later today.

B: When exactly ~~will you call your sister~~?

In (17a), Principle C of binding theory (Chomsky, 1986) prevents the ellipsis clause from containing the proper noun *Alex*, causing a mismatch between the elided pronoun *him* and its antecedent *Alex*, but the utterance is nonetheless acceptable. Similarly, (17b) is acceptable despite the mismatch in indexical pronouns, which is caused by the fact that the antecedent and ellipsis clauses are uttered by different people. This type of mismatch was termed "vehicle change" by Fiengo and May (1994), who proposed that the IDENTITY condition that governs ellipsis must be defined in terms of equivalence classes such that mismatches between elements in the same equivalence class are permitted. Specifically, they argue that lexical items that are co-referential, as is the case in both (17a) and (17b), are to be considered equivalent in the relevant sense, which allows them to maintain a syntactic IDENTITY condition for ellipsis: elided material must be syntactically identical to its antecedent *modulo* "vehicle change."

This is the first example of what I mean by the term "fine-tuning": by adding qualifications to the definition of IDENTITY, Fiengo and May succeed in "capturing" a wider range of data, but those qualifications make the theory less parsimonious because they are not independently motivated beyond the very data they are designed to capture. In Merchant's (2001, p. 25) words, "[t]o pursue a theory of [syntactic IDENTITY] while considering the cases of 'vehicle change' to have been sufficiently dealt with simply by naming them is to confuse the diagnosis with the cure."

To avoid this issue, Merchant (2001) proposes a more radical revision of the IDENTITY condition: On the basis of examples like (17) (among several other types of examples), he

rejects the notion that ellipsis is governed by *syntactic* IDENTITY altogether. In its place, he proposes a *semantic* IDENTITY condition, known as "e-GIVENness," which remains one of the most influential theories of ellipsis to date. At its core, e-GIVENness is based on Schwarzschild's (1999) notion of Givenness, which was designed to account for patterns of focus marking and accent placement:

(18) **Givenness** (Schwarzschild, 1999): An utterance U counts as GIVEN iff it has a salient antecedent A and modulo ∃-type shifting, A entails the Existential F-Closure of U.[3]

For example, in (19) the Verb Phrase *eat a green apple* is GIVEN because its existential F-closure—i.e. the result of replacing all focus-marked phrases with existentially quantified variables; in this case: $\exists X[X$ate a green apple$]$—is entailed by the first clause.

(19) John ate a green apple but Max didn't eat a green apple.

In addition to this independently motivated definition of Givenness, however, Merchant's e-GIVENness additionally contains a "reverse entailment" requirement that only applies to ellipsis (hence the "e" in e-GIVENness):

(20) **e-GIVENness** (Merchant, 2001): an expression ε is e-GIVEN iff ε has a salient antecedent A such that, modulo ∃-type shifting, A entails the existential F-closure of ε, and ε entails the existential F-closure of A.

To illustrate how this works, let us apply it to one of the "vehicle change" examples mentioned earlier:

(21) They arrested John even though he didn't think they would ~~arrest him~~. = (17a)

[3] Schwarzschild's final definition of Givenness contains additional qualifications, but for our purposes they are irrelevant and thus omitted for clarity.

a. $[\![A]\!] = arrest(they, John)$

b. $[\![\varepsilon]\!] = arrest(they, John)$

c. F-clo$(A) = \exists x[arrest(they, x)]$ (assuming that *John* is F-marked)

d. F-clo$(\varepsilon) = \exists x[arrest(x, John)]$

Since $[\![A]\!]$ entails F-clo(ε), and $[\![\varepsilon]\!]$ entails F-clo(A), the elided material is e-GIVEN and ellipsis is permitted. Notice that the reliance on (mutual) semantic entailment, rather than syntactic isomorphism, is responsible for ignoring the mismatch between *John* and *him*.

It is crucial to note that the independently motivated part of e-GIVENness by itself—Schwarzschild's notion of Givenness—vastly overgenerates as a constraint on ellipsis: any context entails an unlimited number of propositions, but only those that are "close enough" to the antecedent can be elided. Therefore, the "reverse entailment" clause in the definition of e-GIVENness is crucial, yet there are to my knowledge no other linguistic phenomena that require anything like it: the notion that an antecedent element has to be entailed by the "downstream" element that depends on it is, to my knowledge, unique to e-GIVENness. While this raises concerns of theoretical parsimony, which will be addressed in Section 2.3, the proposal additionally has empirical short-comings in its predictions for both VP-ellipsis and sluicing, which we turn to next.

First, consider the following examples of VP-ellipsis from Hartman (2009):[4]

(22) a. *Mary will beat someone at chess, and John will ~~lose to someone at chess~~, too.

[4]Nash-Webber (1977) makes a very similar observation that predates both Hartman (2009) and Merchant (2001) by more than two decades:

(i) *Bruce sold a waffle iron to Wendy, and an electric wok was ~~bought by Wendy~~ too.
(Nash-Webber, 1977, ex. 22)

This example involves two independent mismatches: a voice mismatch between an active antecedent clause and a passive ellipsis clause; and a lexical mismatch between the relational opposites *sold* and *bought*, which is equivalent to Hartman's examples.

(Hartman, 2009, ex. 3)

b. *John lives with his grandparent, and Bill also does ~~live with his grandchild~~.

(Hartman, 2009, ex. 8)

Both of these interpretations are falsely predicted to be available because they contain relational opposites that ensure that the mutual entailment requirement of e-GIVENness is met: beating someone at chess entails that someone loses (i.e., the antecedent in (22a) entails the existential closure of the elided VP), and having lost to someone entails that someone won (i.e., the ellipsis clause entails the existential closure of the antecedent VP). Likewise, the relational nouns *grandparent* and *grandchild* in (22b) ensure mutual entailment in much the same way (i.e., John living with his grandparent entails that someone, namely his grandparent, lives with their grandchild, and vice versa), but ellipsis is again impossible.

Second, sluicing is unacceptable when the antecedent clause and the ellipsis clause differ in voice (Chung, 2006, 2013; Merchant, 2013b),[5] as illustrated by the following minimal set:

(23) a. Joe was murdered, but we don't know by who ~~he was murdered~~.

b. Someone murdered Joe, but we don't know who ~~murdered him~~.

c. *Joe was murdered, but we don't know who ~~murdered him~~.

d. *Someone murdered Joe, but we don't know by who ~~he was murdered~~.

When the antecedent and ellipsis clauses are both in either active or passive voice, as in (23a-b), sluicing succeeds, but when there is a voice mismatch, as in (23c-d), ellipsis appears to be impossible. Since this difference in voice does not affect the truth-conditional meaning of the utterance, however, e-GIVENness is satisfied throughout and sluicing is predicted to be insensitive to voice mismatches (Merchant, 2013b). This issue (along with other related ones) has motivated a departure from purely semantic IDENTITY conditions towards more elaborate "hybrid" conditions

[5]Chapter 5 will present evidence that complicates this picture, but those qualifications need not concern us here.

that reference both semantic and syntactic representations (Chung, 2006, 2013; Merchant, 2013a, 2013b; AnderBois, 2014; Barros, 2014).

A particularly influential hybrid IDENTITY proposal is due to Chung (2006), who builds on e-GIVENness and extends it by a lexico-syntactic requirement that she refers to as the "No New Words constraint." According to this constraint, which is sometimes referred to as "Chung's generalization," ellipsis is only permitted when the elided material contains no lexical items that are not also contained in the antecedent. By preventing the ellipsis of "New Words" (underlined), Chung's generalization correctly classifies the Hartman cases, voice mismatches, as well as other argument-structure mismatches as ungrammatical:

(24) a. *Mary will beat someone at chess, and John will ~~lose to someone at chess~~. ≈ (23a)

b. *John lives with his grandparent, and Bill also does ~~live with his grandchild~~.
= (23b)

c. *Joe was murdered, but we don't know who ~~murdered him~~.[6] = (24c)

d. *Someone murdered Joe, but we don't know by who ~~he was murdered~~. = (24d)

e. *Mary was flirting, but they wouldn't say who ~~she was flirting with~~.

(Chung, 2006, ex. 19d)

While banning lexical mismatches no doubt improves on previous proposals in terms of empirical coverage, this improvement comes at the cost of exploding complexity because it ends up having to explicitly carve out several exceptions. First, syntactic traces have to be ignored in order to accommodate sluices that involve a trace with no explicit correlate in the antecedent clause (Merchant, 2013a):

(25) John is eating, but I can't see what$_i$ ~~he is eating t_i~~.

[6]Chung (2006) explicitly treats *murdered.*PASSIVE and *murdered.*ACTIVE as non-identical in order to rule out passive-active mismatches of this kind.

This type of sluicing, which Chung et al. (1995) termed "sprouting," accounts for approximately 75% of cases found in natural language corpora (Nykiel, 2010; Anand & Hardt, 2016), and in order to avoid falsely classifying them as ungrammatical, the elided trace must be excepted from the "No New Words" constraint.

Secondly, while e-GIVENness was ostensibly motivated by the desire to render arbitrary "vehicle change" equivalence classes obsolete, Chung's ban against lexical mismatches reintroduces the need for an exception for syntactically distinct but co-referential lexical elements, like *Alex* and *him* in (17a), repeated below.

(26) They arrested Alex, though he thought they wouldn't ~~arrest~~ ~~him~~. = (17a)

Finally, the definition of "vehicle change" must be expanded to cover mismatching elements that are syntactically co-indexed but not co-referential, as shown in the following examples:

(27) a. [Which person]$_1$ will win the next election and by what margin ~~will~~ ~~they$_1$~~ ~~win it~~?
(Ginzburg, 1992, ex. 302a)

b. Who$_1$ did the suspect call t_1 and when ~~did the suspect call~~ ~~them$_1$~~?
(Merchant, 2001, ex. 112b)

c. I don't know who$_1$ said what$_2$ or why ~~they$_1$~~ ~~said it$_2$~~. (Rudin, 2019, ex. 19a)

In these examples, an elided pronoun must count as non-distinct from a wh element in the antecedent clause (or the trace it leaves behind). While Merchant (2001) claims that such examples are covered under e-GIVENness, Rudin (2019) argues, following Merchant (1999), that the two elements must be semantically distinct because the antecedent receives an interrogative interpretation whereas the ellipsis clause is non-interrogative. More specifically, it appears that the elided pronoun refers to an implicit answer to the question raised by the antecedent, and as

long as questions are semantically distinct from their answers, mutual entailment should fail in those cases.

Despite its return to "vehicle change" and the need to carve out additional *ad-hoc* exceptions, some version of Chung's generalization has since been incorporated into various other "hybrid IDENTITY" proposals (AnderBois, 2010; Chung, 2013; Merchant, 2013a, 2013b; AnderBois, 2014, among others).[7] Furthermore, a similar hybrid approach, which combines syntactic and semantic IDENTITY by way of a syntactic inference mechanism, has been proposed by Thoms (2015). Following Fox (1999), Thoms proposes to extend the definition of IDENTITY by an inferential mechanism termed "antecedent accommodation," whereby IDENTITY is evaluated against an augmented set of potential antecedents, containing both the explicitly available one as well as a set of additional, accommodated antecedents that are derived from it. Adapting machinery proposed by Katzir (2007) for generating alternative utterances to be used for deriving implicatures, Thoms (2015, ex. 51) posits the following algorithm for defining the augmented set of potential antecedents.

(28) a. A set of additional antecedents, Ad(A), may be accommodated on the basis of the original (overt) antecedent A.

b. The members of Ad(A) are alternatives derived from A by

(i) deletion

(ii) contraction

(iii) substitution

c. Complexity constraint: all members of Ad(A) must be at most as complex as the

[7] Note that not all hybrid IDENTITY theories adopt e-GIVENness as the semantic component. AnderBois (2014) and Barros (2014), for example, instead propose that the ellipsis clause must address a salient Question Under Discussion (Roberts, 1998, 2012). This distinction will be addressed in more detail in Chapter 6.

overt antecedent A.[8]

d. Semantic constraint: all members of Ad(A) must be semantically identical to the overt antecedent A.

While Thoms' IDENTITY condition itself is purely syntactic in nature, the set of accommodatable antecedents is constrained by semantic IDENTITY. Together, these assumptions amount to a hybrid account that derives Chung's generalization by prohibiting additions or substitutions of semantically distinct lexical material. It thus rules out Hartman cases on the grounds that they involve semantically non-identical lexical mismatches (e.g., *grandparent* vs. *grandchild*), while allowing "vehicle change" substitutions that do not violate semantic IDENTITY (e.g., *Alex* vs. *him*).

Since all of the examples we have seen so far are problematic for either syntactic or semantic IDENTITY theories, hybrid approaches are able to achieve improved empirical coverage by referencing both levels of representation. However, there are several remaining counterexamples in which ellipsis succeeds in spite of various types of mismatches (in finiteness, tense, modality, and polarity) that violate e-GIVENness, the No New Words constraint, or both, and which cannot be rescued by antecedent accommodation, either:

(29) a. The baseball player went public with his desire to be traded. He doesn't care where ~~he will be traded~~. (finiteness mismatch; Rudin, 2019, ex. 21b)

b. Your favorite plant is alive, but you can never be sure how long ~~it will be alive~~. (tense mismatch; Rudin, 2019, ex. 22)

c. Sally knows that there is always the potential for awful things to happen, but she doesn't know when ~~awful things might happen~~.

[8]The purpose of this constraint is to explain various patterns that are immaterial for our present discussion; see Thoms (2015) for details. Note, however, that this constraint incorrectly rules out the "extended vehicle change" example in (27b), since the elided pronoun *they* is strictly more complex (according to Thoms' definition of complexity) than the syntactic trace in the antecedent that constitutes its correlate.

(modality mismatch; Rudin, 2019, ex. 23a)

d. Either the Board grants the license by December 15 or it explains why ~~the Board did not grant the license by December 15~~.

(polarity mismatch; Kroll, 2019, ex. 30)

The underlined lexical elements violate Chung's generalization—as well as Thoms' constraint against additions and semantically non-identical substitutions—and should therefore render each of these examples ungrammatical. Based on examples like these, Rudin (2019) proposes an additional amendment[9] to the IDENTITY condition by restricting its domain. Specifically, he observes that the mismatches in (29) are all located above what he terms the "eventive core" of the elided material (defined as the highest elided *v*P) and concludes that the domain of IDENTITY must be restricted accordingly: only the material inside the eventive core is required to be syntactically identical to the antecedent, while all elements outside of it can be freely elided without being subject to IDENTITY. This is a rather consequential amendment, since it undermines a central intuition that is shared across all theories of ellipsis, i.e. that material can only be elided if it is provided by the context. Since the IDENTITY requirement is an attempt at defining contextual "Givenness" in a way that captures the distribution of ellipsis, restricting it to a proper subset of the elided material leaves the elidability of the exempted material unexplained.

In addition to this theoretical concern and the additional complexity Rudin's proposal introduces compared to previous accounts, it also faces empirical challenges that come in various flavors. First, as pointed out by Kehler (personal communication), Rudin's IDENTITY condition falsely rules out the following example of sluicing, which is modeled after cases of VP-ellipsis from Kehler (2016):

[9] While Rudin (2019) rejects e-GIVENness and instead advocates for a return to a purely syntactic IDENTITY condition, he ends up having to encode the same exceptions discussed above in the context of Chung's generalization: traces in sprouting cases, "vehicle change" of co-referential elements, and mismatches between lexical elements that do not co-refer but are nonetheless syntactically co-indexed are all explicitly excepted from his IDENTITY condition.

(30) [All of the girls]$_1$ hope that they$_1$ will be asked to the prom by someone, but in Sue$_2$'s case, I can't imagine who ~~she$_2$ will be asked by~~.

The elided pronoun *she* is co-indexed with *Sue*, but *Sue* is not part of the antecedent. While *she* intuitively corresponds to *all of the girls*, that correspondence is established by the phrase *in Sue's case* in a way that does not involve syntactic co-indexation (see Kehler, 2016, for an analysis of *as for X* phrases in terms of the Question Under Discussion). Indeed, Rudin (2019) discusses a similar example in footnote 14 and concedes that it is falsely classified as ungrammatical under his IDENTITY condition.

Secondly, consider the following German sluice from Paape (2016) in which the elided material requires a different word order than the antecedent:

(31) Ein Sympathisant der Opposition hatte die Rebellen laut einem Bericht
A sympathizer.NOM of the opposition had the rebels.ACC according to a report
maßgeblich unterstützt, aber die Regierung konnte nicht nachweisen, wie ~~der~~
decisively supported, but the government could not prove, how ~~the~~
~~Sympathisant der Opposition die Rebellen unterstützt hatte~~.
~~smypthesizer of the opposition the rebels.ACC supported had~~.

'According to a report, a sympathizer of the opposition had supported the rebels, but the government couldn't prove how.'

German main clauses require the finite verb to be in second position, as is the case in the antecedent clause in this example. Subordinate clauses (like the sluiced clause), however, require verb-final word order, which causes a word-order mismatch between the elided material and the antecedent clause. Since Rudin (2019) defines syntactic IDENTITY in terms of a "structure matching" algorithm that requires all elided syntactic heads to be dominated by the exact same sequence of syntactic nodes as their correlates in the antecedent, it is sensitive to word order

mismatches and falsely classifies this fully acceptable example as ungrammatical.[10]

Finally, Rudin's proposal is—just like all other existing IDENTITY theories that I am aware of—vulnerable to a set of lexical mismatches that are not covered under vehicle change, syntactic co-indexing, or Chung's exception for inflectional mismatches. Consider first the following examples from Kehler (2002a):

(32) a. Mary's boyfriend gave her his school picture, just as all schoolboys do ~~give their girlfriends their schoolpicture~~.

b. A: Bob's mother cleans up after him all the time.
B: I'm surprised; most parents these days won't ~~clean up after their children~~.

In these examples there is a lexical mismatch between the elided object NPs *girlfriends* and *children* and their correlates *boyfriend* and *him*, respectively. While these examples are superficially similar to the Hartman cases discussed above in that they involve relational nouns, they afford the opposite conclusion: while Hartman's examples demonstrate that e-GIVENness generates interpretations that do not, in fact, arise, the examples here highlight an *undergeneration* problem with lexico-syntactic IDENTITY by demonstrating that interpretations that shouldn't arise, do. Similarly, the following two examples, which are representative of cases that will be discussed in more detail in Chapters 4 and 6, respectively, involve mismatching main verbs but are nonetheless relatively acceptable:

(33) a. A: Can I borrow your textbook over the weekend?
B: I can't ~~lend it to you~~: I'll need it myself.

b. Fan: Can I please get a few autographs?
Manager: Sure, how many ~~do you want~~?

[10] See Merchant (2001, p. 21) for a similar argument regarding Dutch, another V2 language.

All of the lexical mismatches in (32) and (33) are located within the "eventive core" and thus violate Rudin's IDENTITY condition.[11]

Up to this point, we have seen various amendments to the basic idea that the elided material must be identical to its antecedent in order for ellipsis to be grammatical. As summarized in Table 2.1 each amendment was motivated by otherwise problematic ellipsis data and has led to an incremental "fine-tuning" of the IDENTITY condition in order to maximize its empirical coverage. Despite this decades-long "quest for identity" (Lipták, 2015, p. 155), numerous remaining mismatches elude even the most complex IDENTITY theories to date, giving rise to what I refer to as the "IDENTITY crisis." In the next two sections, I will describe two other forms of "fine-tuning" that contribute to this crisis by targeting theory-external moving parts. First, I will briefly discuss a strategy for dealing with mismatch data that involves reanalyzing mismatching elements, rather than fine-tuning the IDENTITY condition itself, with the goal of construing them as "underlyingly identical." I will illustrate this strategy with two examples: mismatching polarity items; and "category mismatches" that result from VP-ellipsis or sluicing with nominal antecedents.

2.1.2 Exploiting other moving parts: reanalyzing mismatching elements

Consider first the following examples of VP-ellipsis, which involve mismatching polarity items (~~strike-out~~ annotation and underlining are mine):

(34) a. John doesn't see anyone, but Bill does ~~see~~ ~~someone~~. (Sag, 1976, ex. 2.3.39)

b. John saw someone, but Mary didn't ~~see~~ ~~anyone~~. (Merchant, 2013a, ex. 2)

c. We haven't decided to blacklist any firms. But there's a chance we might ~~blacklist~~ ~~some~~ ~~firms~~. (Hardt, 1993, ex. 68)

[11] While Rudin (2019) is primarily concerned with sluicing, he does suggest that his IDENTITY condition may be extended to VP-ellipsis, in which the eventive core necessarily contains all of the elided material. Even if his proposal were restricted to sluicing, however, the challenge raised by (33b) remains.

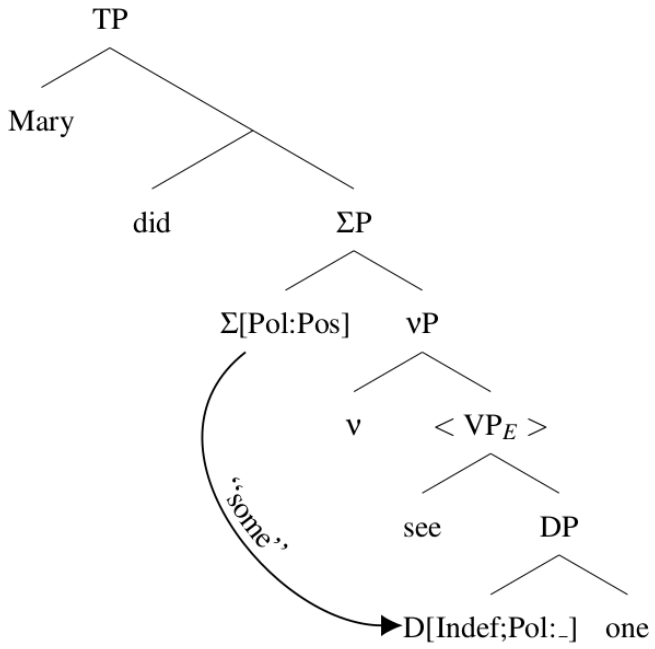
Table 2.1: Cross-tabulation of IDENTITY proposals and mismatch phenomena. ✓s indicate that phenomenon and theory are consistent.

Phenomenon / type of mismatch		Exact IDENTITY	e-GIVENness	Chung (2006) etc.	Rudin (2019)
close correspondence with antecedent	(15)	✓	✓	✓	✓
classic "vehicle change"	(17)		✓	✓	✓
Hartman cases	(22)	✓		✓	✓
voice-mismatched sluicing	(23)	✓		✓	✓
elided traces with no correlate	(25)		✓	✓	✓
wh/answer "vehicle change"	(27)				✓
finiteness, modality, tense, polarity mismatches	(29)				✓
"In *X*'s case" examples	(30)				
word-order mismatches	(31)				
mismatching relational nouns	(32)				
main-verb mismatches	(33)				

d. I could find no solution, but Holly might ~~find a solution~~. (Johnson, 2001, ex. 107)

Merchant (2013a) suggests that the mismatching polarity items in each of those cases can be analyzed as "underlyingly identical" to their correlates in the antecedent. According to his analysis, polarity items are abstract elements at the level of syntax and lexicalized in agreement with a "polarity phrase"—ΣP—that is located above the ellipsis site and therefore not subject to IDENTITY. Indeed, if the IDENTITY condition is defined over syntactic representations and polarity items are modeled as identical at that level of representation by analyzing all differences between them as the result of a post-syntactic lexicalization process, they do not, in fact, violate syntactic IDENTITY:

(35) (see Merchant, 2013a, ex. 3)



Similar analyses in which the mismatching elements are raised out of the domain of IDENTITY have been proposed for other types of mismatches in VP-ellipsis as well, including inflectional mismatches (see e.g., Lasnik, 1995; Lipták, 2015) and voice mismatches (Merchant, 2013b).[12] As far as polarity items are concerned, Merchant (2013a) argues that there are independent reasons for analyzing them in this way, but he also concedes that restricting oneself to a certain analysis of non-elliptical elements in the service of salvaging an otherwise empirically inadequate theory of ellipsis could serve as "a *reductio ad absurdum* of the syntactic approach to the elliptical identity question" (p. 459). I will discuss theoretical considerations that support this conclusion in Section 2.3, but for the time being the crucial point is simply that the "fine-tuning approach" to IDENTITY theories of ellipsis does not merely target theory-internal parameters, such as the definition of IDENTITY itself, but may also exploit degrees of freedom inherent in the way other linguistic elements are analyzed.

Another example of analyzing mismatching elements as "underlyingly identical" involves

[12]The empirical picture with respect to voice-mismatched VP-ellipsis is, in fact, much more complex. While Merchant (2013b) predicts that voice mismatches should be categorically acceptable, many cases are heavily marked and the overall picture is one of gradient acceptability (Kehler, 2000; Arregui, Clifton, Frazier, & Moulton, 2006; Kertz, 2013; Poppels & Kehler, 2019). Chapter 3 will go into more detail on this issue.

category mismatches, i.e. cases in which the elided material and its antecedent belong to different syntactic categories. Consider the following examples from Kehler (1993b, attributed to Gregory Ward) and Hardt (1993):

(36) a. The letter deserves a response, but before you do ~~respond~~, ...
(Kehler, 1993b, ex. 20)

b. Meanwhile, they sense a drop in visitors to the city. Those who do ~~visit the city~~, they say, are not taking cabs. (Kehler, 1993b, ex. 21)

c. Harry used to be a great speaker, but he can't ~~speak~~ anymore, because he has lost his voice. (Hardt, 1993, ex. 114)

In each of these examples, the VP-ellipsis site appears to depend on an NP as its antecedent (*response*, *visitors*, and *speaker*, respectively), which is at odds with the IDENTITY requirement regardless of whether it is defined over syntactic or semantic representations. To address this issue, Johnson (2001), following Fu, Roeper, and Borer (2001), suggests that antecedent NPs can be analyzed as being derived from underlying VP representations, and that those VPs are available to serve as antecedents to the ellipsis site. This analysis correctly predicts that transparently deverbal nouns, such as *robber* in (37a), should be more acceptable as antecedents for VP-ellipsis than morphologically opaque ones, such as *thief* in (37b) (Merchant, 2013a). However, other minimal pairs that produce similar contrasts cannot be explained in this way: the same nominalizations that were felicitous antecedents in (36) do not work in (38) (see Kehler, 1993b, 2000, for a coherence-based explanation for this contrast).[13]

(37) a. That man is a robber, and when he does ~~rob places~~, he tries not to make any noise.

b. #That man is a thief, and when he does ~~steal stuff~~, he tries not to make any noise.

[13]For many more examples of nominal-antecedent VP-ellipsis, both felicitous and infelicitous, along with insightful theoretical discussion, see Miller and Hemforth (2014).

(38) a. #The letter provoked a response from Bush, and Clinton did ~~respond~~, too.

(Kehler, 1993b, ex. 22)

b. #There is a rise in American visitors to the city, and Canadians do ~~visit~~, too.

(Kehler, 1993b, ex. 23)

It is important to note that while the decision to analyze NPs as underlyingly containing a VP is motivated solely based on ellipsis data, doing so does have consequences that fall outside the scope of the theory of ellipsis. For example, if nominalizations underlyingly contain the VPs they derive from, one would expect them to be modifiable by adverbs. Indeed, Fu et al. (2001) cite examples like (39) in support of this prediction, which they judge to be grammatical, but I agree with Kehler (2019) that they are marginally acceptable at best, which undermines Fu et al.'s analysis.

(39) a. #Kim's explanation of the problem to the tenants thoroughly did not prevent a riot. [where *thoroughly* is interpreted as modifying *explanation*]

(Fu et al., 2001, ex. 1a)

b. #The occurrence of the accident suddenly disqualified her. [where *suddenly* is interpreted as modifying *occurrence*] (Fu et al., 2001, ex. 1b)

The picture becomes even more complicated once we consider the fact that NPs can not only serve as antecedents to VP-ellipsis, but also to sluicing. In a corpus study, Beecher (2007) found the following naturally occurring examples in which the sluicing site depends on the underlined nominal antecedent.

(40) a. We're on to the <u>semi-finals</u>, though I don't know who against.

(Beecher, 2007, ex. 8a)

b. The only thing I can come up with is <u>contamination</u> but I do not know what from.

(Beecher, 2007, ex. 10a)

Since sluicing involves an elided clause, rather than merely a VP, we would have to extend Johnson's (2001) analysis: rather than deriving from VPs, the NPs in question have to be analyzed as underlyingly containing full clauses that are capable of providing antecedents for both sluicing and VP-ellipsis. Even more concerning is the fact that the NP does not provide any morphological clues as to what the underlying clause must be, which—by analogy with the *thief* example above—suggests that ellipsis should be infelicitous. We will revisit this issue in greater detail in Chapter 6, which presents experimental evidence on nominal-antecedent sluicing. For our present purpose, it is sufficient to emphasize that the analysis of NPs as containing covert VPs or clauses represents another example of "fine-tuning" theory-external parameters in order to maintain the assumption that ellipsis is governed by an IDENTITY relation in light of evidence to the contrary. The next section describes a third strategy of bringing mismatch data into alignment with IDENTITY theories, which involves complementing the grammatical constraints imposed by the IDENTITY condition with a processing theory that accounts for cases of non-identity as "acceptable ungrammaticality" (Arregui et al., 2006; Frazier, 2013). I will briefly sketch the main ideas behind this approach below, and Chapter 3 will evaluate this approach in more detail.

2.1.3 Exploiting degrees of freedom in the grammar/processor partition

At least since Miller and Chomsky's (1963) discussion of center embedding, there has been wide-spread agreement that the acceptability of utterances is affected both by grammatical constraints as well as processing factors. This creates an ambiguity: looking at any particular sentence, it is non-trivial to determine whether its status is to be attributed to the grammar, aspects of the processing system, or both. In an influential line of research, Lyn Frazier and her colleagues exploit this ambiguity by arguing that elliptical utterances with syntactically mismatched antecedents that are more acceptable than predicted under IDENTITY can be explained

in terms of processing factors. According to their proposal, which has been known since Arregui et al. (2006) as the "Recycling Hypothesis," elliptical utterances with non-identical antecedents trigger the same processing mechanism that is responsible for the syntactic reanalysis of garden-path sentences, such as the one in (41).

(41) The horse raced past the barn fell.

In such sentences, comprehenders tend to initially adopt a parse of the sentence that ends up being inconsistent with later parts of the sentence; in this case, the past-participle *raced* is initially parsed as the main verb, which is inconsistent with encountering another main verb in *fell*. Once the inconsistency is detected, the processor has to reanalyze the preceding linguistic material in order to identify the correct parse. While some garden-path sentences are more difficult to recover from than others (the one in (41) is notoriously difficult), the fact that comprehenders can recover from them at all is a testament to the fact that the processor is capable of reanalyzing past linguistic material that is inconsistent with its current input.

According to the Recycling Hypothesis, the same mechanism is triggered when the parser is confronted with an elliptical utterance whose antecedent violates IDENTITY, and the gradience in acceptability that is associated with various types of mismatches follows, by hypothesis, from two factors: first, the amount of "recycling" work that is required to turn a "flawed" antecedent into one that obeys the IDENTITY requirement; and secondly, the amount of evidence the Recycler has about the operations that are required. To illustrate, consider the following examples from Arregui et al. (2006):

(42) a. Singing the arias tomorrow night will be difficult, but Maria will ~~sing the arias tomorrow night~~.

b. Singing the arias slowly tomorrow night will be difficult, but Maria will ~~sing the arias slowly tomorrow night~~.

c. Tomorrow night's singing of the arias will be difficult, but Maria will ~~sing the arias tomorrow night.~~

d. Tomorrow night's slow singing of the arias will be difficult, but Maria will ~~sing the arias slowly tomorrow night.~~

The antecedent clauses in (42a-b) contain a verbal gerund, while (42c-d) contain a nominal gerund. While both verbal and nominal gerunds behave syntactically as NPs (e.g., both act as subjects), Arregui et al. assume that verbal gerunds, but not nominal ones, contain an embedded VP, as evidenced by the fact that the former can be modified by an adverb, as in (42b), whereas the latter require an adjectival modifier, as in (42d). Given these assumptions, the Recycling Hypothesis makes two predictions about the acceptability of ellipsis. First, nominal-gerund antecedents should be less acceptable because they require more recycling work in order to satisfy IDENTITY.[14] Secondly, the presence of the adjectival modifier in (42d) should further decrease acceptability because turning it into an adverb to match the elided VP presents the Recycler with additional work. Across two experiments, Arregui et al. show that the first prediction is consistently borne out: the examples involving verbal-gerund antecedents were consistently rated as more acceptable than their nominal-gerund counterparts. The second prediction, however, was not borne out: the presence of a modifier did not significantly affect acceptability in either the first experiment or the second one.[15]

While the authors attribute the lack of evidence for the predicted effect of the modifiers to a lack of statistical power, it does cast doubt over the empirical validity of Recycling Hypothesis. Along a similar vein, Chapter 3 describes a series of experiments that are inconsistent with a number of other predictions from the Recycling Hypothesis. For now, however, my goal is

[14] While the verbal-gerund cases do provide a suitable antecedent VP, Arregui et al. suggest that locating a VP that is embedded under an NP in subject position nonetheless requires extra processing effort compared to VP antecedents in "canonical" syntactic positions.

[15] Note that similar arguments have been put forth for a range of other mismatch phenomena, including voice mismatches (Arregui et al., 2006; Grant, Clifton Jr, & Frazier, 2012).

merely to emphasize that the strategy of attributing gradient acceptability patterns that appear to violate IDENTITY to processing factors reflects yet another way of "fine-tuning" theory-external parameters in order to insulate IDENTITY theories from seemingly contradictory evidence. The key question is whether or not the parameters in question are independently motivated. Indeed, at an abstract level the Recycling Hypothesis is motivated by an analogy with a processing mechanism that is independently needed to explain the possibility of garden-path recovery. However, this analogy does not extend to the level of detail that would be required in order to make independently motivated predictions. For example, to my knowledge the set of operations the Recycler is assumed to be capable of and the relative processing burden they incur has never been specified explicitly, let alone justified in terms of garden-path recovery. Likewise, various aspects of the linguistic context—including presuppositions (Arregui et al., 2006), the salience of a relevant Question Under Discussion (Grant et al., 2012), among others—have been said to count as "evidence" that facilitates the Recycling process, but these stipulations are not independently motivated by non-elliptical phenomena. In the absence of independent constraints on the theory-external parameters the Recycling Hypothesis exploits, it therefore risks "confusing the diagnosis with the cure," to use Merchant's (2001) phrase.

2.1.4 Summary

In the preceding sections, I discussed three types of "fine-tuning" strategies that have been pursued with the goal of accommodating otherwise problematic evidence under the assumption that ellipsis is governed by IDENTITY: fine-tuning the definition of IDENTITY itself; re-analyzing mismatching elements as "underlyingly identical;" and construing certain mismatches as cases of "acceptable ungrammaticality" that reflect properties of the processor, rather than the grammar. I highlighted five sources of "degrees of freedom" that these strategies exploit: (i) the level of representation at which IDENTITY is defined (we have seen syntactic, semantic, and hybrid conditions); (ii) ad-hoc exceptions, for example for traces or mismatches that result from "vehicle

change;" (iii) the domain in which IDENTITY applies (e.g. Rudin's "eventive core" restriction); (iv) theory-external assumptions about the underlying representation of mismatching elements; and (v) details about the grammar/processing partition. The "fine tuning" approach is problematic for two reasons: first, since the incremental fine-tuning of both theory-internal and -external parameters is tailored to ellipsis-specific observation and not independently motivated, it risks "overfitting" the theory of ellipsis to the data, which undermines its explanatory value. Second, following H&S (Hankamer & Sag, 1976; Sag & Hankamer, 1984), IDENTITY theories are based on the fundamental assumption that ellipsis is governed by a *sui generis* mechanism that does not explain any other phenomena outside the domain of ellipsis, which raises concerns of theoretical parsimony. Neither of these concerns apply to referential theories of ellipsis, which I turn to next.

2.2 Referential theories of ellipsis

In this section, I will review a class of theories that reject the notion that ellipsis is governed by an IDENTITY condition and instead analyze it as a form of discourse reference, according to the following basic assumption (Webber, 1978; Hardt, 1993; Kehler, 1993a, 1993b, 2000; Barker, 2013; Poppels & Kehler, 2019, and many others).

(43) **Central claim:** Elliptical utterances contain a silent *pro*-form that completes the meaning of the ellipsis clause anaphorically by recruiting the same mechanism that governs non-elliptical forms of discourse reference.[16]

Recall that H&S (Hankamer & Sag, 1976; Sag & Hankamer, 1984) argued that ellipsis does

[16]Following Klein (1987), several referential theories of ellipsis have been formalized within the framework of Discourse Representation Theory (e.g., Hardt, 1992; Bos, 1993). By contrast, I remain agnostic as to the exact nature of the mechanisms that enable discourse reference. Many aspects of discourse reference are poorly understood, and the processes and constraints that enable inferential reference resolution remain particularly mysterious to date. Since those aspects are of central importance to the work presented here, adopting any particular implementational framework is unlikely to yield any theoretical insights. Instead, I propose deriving predictions of referential theories of ellipsis by analogy with non-elliptical forms of discourse reference.

not engage the discourse-referential system, and instead is governed by an ellipsis-specific dependency between the ellipsis site and its antecedent. They came to this conclusion based on the observation that ellipsis differs from non-elliptical forms of discourse reference in two fundamental ways: its inability to refer exophorically to situationally evoked entities; and its sensitivity to morphosyntactic properties of the antecedent, such as voice.

In this section, I will revisit both of these arguments against a larger set of empirical data and argue that their conclusion is not warranted. In Section 2.2.2, I will first review the evidence regarding exophoric ellipsis and extend Hankamer and Sag's reasoning to five other diagnostic properties of discourse reference. As we will see, both VP-ellipsis and sluicing pattern closely with other discourse-referential devices. Section 2.2.3 then addresses the issue of sensitivity to the morphosyntactic form of the antecedent, reviewing evidence that, contra H&S, non-elliptical forms of reference do exhibit similar behavior. This will prepare the ground for asking what factors affect the acceptability of using particular referring expressions in particular contexts and outlining open questions about the distribution of ellipsis from the perspective of referential theories. I will begin by summarizing key properties of non-elliptical forms of discourse reference.

2.2.1 What is discourse reference?

The following utterances all contain at least one discourse-referential expression (underlined):

(44) a. Robert tried to address the room, but nobody was listening to <u>him</u>.

b. [Context: Approaching someone else's dog in the street.]
Can I pet <u>her</u>?

c. [Context: Walking up to a balloon salesperson]
How much is the blue <u>one</u> with green stripes?

(adapted from Nash-Webber, 1977, ex. 11)

d. Susan called Becky to discuss their dinner plans.

e. Harry threw up and Sam stepped in it.

(Prince, 1981, ex. 13b; based on Tic Douloureux, 1971)

f. When Jack was kidnapped, they kept him in a dark room for days.

(adapted from similar examples in Clark, 1975)

At least since Karttunen (1976), it has been widely assumed that referring expressions like those in (44) do not refer to entities in the world directly, but rather to representations of them in a shared representational space between interlocutors, known as the "discourse model." Each interlocutors' model of the discourse contains a repository of discourse entities that they mutually know to be available for reference.[17] Discourse entities can be introduced into the discourse model in three ways. First, they can be introduced explicitly by their linguistic antecedent, as, for example, in (44a) in which the definite pronoun *him* refers to the person denoted by the antecedent NP *Robert*. Secondly, referents can enter into the discourse model directly from the situational context the interlocutors are mutually aware of, without being mediated by a linguistic utterance, as illustrated by the pronoun *her* in (44b) and the indefinite pronoun *one* in (44c). Finally, discourse entities can be introduced inferentially, and these inferences can take a variety of forms. For example, *their* in (44d) refers to a set of two individuals that can be inferred by combining the denotations of the two separate antecedent NPs *Susan* and *Becky*. Similarly, the referent of *it* in (44e) is not introduced explicitly and instead must be inferred as the expected product of an explicitly denoted event. Finally, *they* in (44f) succeeds in referring to Jack's kidnappers although they are not mentioned explicitly and must instead be inferred from the mention of the kidnapping event.

In spite of what the term "discourse entity" might suggest, referential expressions are not

[17] *Mutual* knowledge goes beyond merely *shared* knowledge in that it is recursive (Lewis, 1969; Clark & Marshall, 1981): Interlocutors *A* and *B* mutually know *P* iff *Q*, where *Q* is true iff both *A* and *B* know *P* and *Q*. Mutually known discourse referents are thus entities that both interlocutors know to be available for reference and for which they both know that they both know that they are potential referents, and so forth.

restricted to individuals or sets of individuals: interlocutors can also refer to a range of other types of objects, including propositions, as in (45a-b), speech acts, as in (45c), situations evoked by potentially large stretches of discourse, as in (45d), as well as events, as illustrated in (45e).

(45) a. A: I read today that coffee can have both positive and negative health implications.
B: I didn't know that.

b. You're not going to believe it, but I just won the Powerball.

c. Donald: I'm going to release my tax returns once the audit is complete.
Everyone: That's a lie.

d. [Context: Alice just finished telling Kendrick an elaborate story about all the frustrating things that happened to her this week.]
Kendrick: That sucks!

e. There's a long history in the US of abusing scripture to advance the causes of bigotry & discrimination. Slaveholders did it. Segregationists did it. White supremacists do it. And it continues. Yet if Christ repeated himself today, they'd likely denounce him as a radical, too. [18]

Given the range of objects that can serve as potential referents, and the fact that even novel entities can become available for reference through inference, comprehenders are faced with the difficult task of identifying the intended referent whenever they encounter a referring expression. Fortunately for them, different referring expressions come with "instructions" that constrain the set of possible referents in a variety of ways. For example, entity-level pronouns that are marked for gender or number, such as *he*, *she*, *they* or *it*, are generally restricted to referents with the same gender/number properties. Those constraints can be observed with respect to entities introduced by conventionally gendered proper nouns, as in (46a), grammatically gender-marked antecedent

[18]This example is from a tweet by U.S. Congresswoman Alexandria Ocasio-Cortez: https://twitter.com/AOC/status/1233795153585897473?s=20.

nouns, as in (46b), and even *pluralia tantum* antecedents, as in (46c), which are grammatically marked as plural even though they denote notionally singular entities. While these constraints associated with grammatical gender/number marking may appear to be due to an agreement relation between the referring expression and its antecedent, they persist even when there is no overt linguistic antecedent and the intended referent is instead evoked situationally, as shown in (47).

(46) a. Have you seen Beth?—Nope, I haven't seen {her | #him} in days.[19]

b. Ich brauche einen neuen Computer, {dieser.MASC | #diese.FEM} hier
I need a new computer.MASC, {this-one.MASC | #this-one.FEM} here
ist zu langsam.
is too slow.
'I need a new computer, this one is too slow.'

c. I haven't worn these pants in years, I even forgot {they | #it} existed.

(47) a. [Context: after rummaging through a pile of laundry for minutes, the speaker triumphantly pulls out the pants he was looking for and announces...]
I found {them | #it}!

b. [Context: after scanning the night sky for the north star, the speaker finally points at it and says...]

(i) Da ist {er.MASC | #sie.FEM}, ich hab {ihn.MASC | #sie.FEM}
There is {he.MASC | #she.FEM}, I have {him.MASC | #her.FEM}
gefunden.
found.
'There it is, I found it.'

[19]Of course referring to Beth as *him* could be perfectly felicitous if Beth was mutually known to identify as masculine.

While constraints based on the gender/number marking of the referring expression may be relatively straightforward, other constraints are less transparent. For example, consider once more (45d), repeated below, in which Kendrick refers to a complex situation described in the preceding discourse. In this context, *that* is perfectly felicitous, but using *it* instead is not. Once the situation has been referred to by *that*, however, the pattern flips: it can now be felicitously referred to by *it*, and *that* is now marked.

(48) [Context: Alice just finished telling Kendrick an elaborate story about all the frustrating things that happened to her this week.]
Kendrick: {That | #It} sucks!
Alice: {#That | It} really does.

Some research suggests that such differences between referring expressions can be cached out in terms of the contextual salience of the intended referent (e.g., Gundel, Hedberg, & Zacharski, 1993; Miller, 2011), its cognitive accessibility (Ariel, 1988, 1991), or the complexity of the referent (Brown-Schmidt, Byron, & Tanenhaus, 2005). As we will see in more detail in Section 2.2.3, these notions remain poorly understood and do not fully capture the constraints on the use of discourse reference. For present purposes, however, it is sufficient to highlight two important points with respect to the differences we observe in (48) and (47): first, referring expressions vary widely with respect to the types of discourse entities they can refer to and the constraints that govern their felicitous use; second, these constraints range from relatively transparent rules about gender/number features to more "nebulous" notions, such as salience, accessibility, referent complexity, and others.

The assumptions about discourse reference outlined above corresponds closely to Sag and Hankamer's (1984) characterization of model-interpretive ("deep") anaphora. I will now return to the central claim behind referential theories of ellipsis and argue that both VP-ellipsis and sluicing exhibit a series of diagnostic properties of discourse reference.

2.2.2 Diagnostic properties of discourse reference

This section discusses six diagnostic properties of discourse reference and argues that both VP-ellipsis and sluicing exhibit all of them. This line of argumentation is an extension of the argument from exophora proposed by H&S (Hankamer & Sag, 1976; Sag & Hankamer, 1984): they identified the capacity for exophoric (i.e., antecedent-less) reference as a diagnostic property shared among forms of discourse reference and then argued that exophoric ellipsis is not felicitous in the same contexts that non-elliptical reference is. I will begin by revisiting the data around exophora and argue for the opposite conclusion, i.e. that exophoric ellipsis is, in fact, possible and that it furthermore appears to be constrained in much the same way as non-elliptical exophora. I will then consider five other diagnostic properties that further support the analogy between ellipsis and discourse reference: the possibility of (i) multiple "split" antecedents, (ii) non-local antecedents, and (iii) cataphoric reference; the ability to (iv) trigger "sloppy" interpretations, and (v) refer to inferentially introduced discourse entities. As we will see, both VP-ellipsis and sluicing exhibit all of these properties, which—following the same logic as H&S—supports the conclusion that they engage the same system that governs other forms of discourse reference. Besides this "argument-by-analogy," some of the data we consider below raise independent challenges for IDENTITY theories, further strengthening the argument that ellipsis should be analyzed as a form of discourse reference. Finally, I will briefly consider other types of ellipsis beyond VP-ellipsis and sluicing at the end of this section.

Exophoric ellipsis

Recall from Section 2.2.1 that discourse entities can be evoked by the situational context in the absence of a linguistic antecedent:

(49) a. [Context: Approaching someone else's dog in the street.]
Can I pet her? = (44b)

b. [Context: A bends down to lift a 500 lb. barbell.]

B: With your back, do you think you should do it?

(Sproat and Ward (1987); cited in Ward, Sproat, and McKoon, 1991, ex. 27b)

According to H&S, exophoric uses of ellipsis are infelicitous, as demonstrated by examples like the following.

(50) [Context: Hankamer produces a gun, points it offstage and fires, whereupon a scream is heard]

a. Sag: Jesus, I wonder who #(has been shot). (Hankamer & Sag, 1976, ex. 43)

b. Sag: Jorge, you shouldn't have #(fired the gun). (Sag & Hankamer, 1984, ex. 5d)

However, the picture is more complex than that. As noted in a footnote in Hankamer and Sag (1976) and later developed in Schachter (1977) in more detail, certain uses of ellipsis do succeed exophorically, as shown in (51). While Hankamer (1978) convincingly shows that many such cases are only felicitous in "illocutionarily charged" utterances (e.g., imperatives) and argues that they are therefore peripheral to the theory of ellipsis, exophoric uses of ellipsis also occur in purely assertive or information-seeking utterances and can be felicitous given sufficient contextual support, as exemplified in (52).

(51) a. [Context: Hankamer brandishes a cleaver, advances on Sag]

Sag: Don't (stab me)! My God, please don't (stab me)!

(Hankamer & Sag, 1976, footnote 18)

b. [Context: John pours another martini for Mary.]

Mary: I really shouldn't (have another martini). (Schachter, 1977, ex. 4)

c. [John comes up to the table where Mary is sitting, makes as if to take one of the spare chairs there]

John: May I (sit)?

Mary: Please do (sit). (Schachter, 1977, ex. 7)

d. [Context: Pouring someone a drink.]

Tell me when (to stop pouring).

(52) a. [Context: soccer commentator describing an attack culminating in a shot on goal, but it isn't clear for a few seconds whether the ball will hit the target]

Henry is through. Valdés comes. Will it (go in)? Will it (go in)? Will it (go in)? Will it (go in)? Yes, it will (go in).[20]

b. [Context: New York Governor Andrew Cuomo in conversation with Chris Hayes on the topic of U.S. concentration camps for migrants.]

They [=the U.S. federal government] don't even want to tell the state how many (children have been detained) and in what facilities (they have been detained). That's why we started the law suit.[21]

c. [Context: I'm ordering in barely fluent French at a bakery in Paris. The counterperson switches to English and asks:]

Which country (are you from)?

Based on similar observations in a corpus analysis of exophoric VP-ellipsis, Miller and Pullum (2013) argue what matters is not whether an elliptical expression is conventionalized or not, but rather whether its discourse conditions are met by the situational context. Following Kertz (2013), they consider two different uses of VP-ellipsis that differ with respect to the information structure of the ellipsis clause and argue that they are felicitous under different discourse conditions: auxiliary-focus VP-ellipsis, in which the subject of the ellipsis clause is deaccented (and often pronominalized); and subject-focus VP-ellipsis, in which the subject

[20]This example naturally occurred during the commentary of a 2009 *Clásico* between Real Madrid and FC Barcelona. At the time of writing, it was available at https://youtu.be/RXeoU4K8UpY?t=357.

[21]This example was observed on the U.S. cable news show *All in with Chris Hayes* on June 21, 2018.

receives contrastive focus. According to Miller and Pullum, VP-ellipsis is felicitous only to the extent that the linguistic or non-linguistic context raises a question that fits the information structure of the ellipsis clause: a polar question for auxiliary-focus VP-ellipsis; and a wh-question for subject-focus VP-ellipsis.[22] Looking back at the example involving VP-ellipsis in (52a), this generalization holds: The non-linguistic context raises a highly salient question as to whether the ball will hit the target, and the commentator's utterance *Will it?* has exactly the information structure Miller and Pullum would predict: auxiliary focus and a pronominalized and deaccented subject. In their corpus investigation, Miller and Pullum (2013) find that naturally occurring examples of exophoric VP-ellipsis overwhelmingly exhibit auxiliary focus, which they attribute to the fact that polar questions arise more easily from the non-linguistic context than wh-questions. The same line of reasoning also explains why even auxiliary-focus VP-ellipsis is only rarely used exophorically: since overt assertions represent an effective way of raising a polar QUD[23] and also provide an explicit antecedent VP, the discourse conditions of (auxiliary-focus) VP-ellipsis are most often met when a linguistic antecedent is present. On that view, then, the "requirement" for an explicit linguistic antecedent is purely epiphenomenal: even non-conventionalized elliptical utterances are felicitous as long as the non-linguistic context provides enough support to satisfy the discourse conditions of the elliptical device in question.

While Miller and Pullum's analysis is restricted to VP-ellipsis and further research will be necessary to develop similar analyses for other types of ellipsis, it is important to note that even non-elliptical forms of discourse reference require extra contextual support to be used exophorically, as illustrated in (53).

[22] Although Miller and Pullum (2013) do not couch their analysis in terms of the Question Under Discussion (QUD; Roberts, 1998; Ginzburg & Sag, 2000; Roberts, 2012) explicitly, the discourse conditions they articulate can be derived from a general principle known as "Question-Answer congruence" (Roberts, 1998, 2012): any utterance is interpreted as the answer to a (often implicit) QUD with the same information structure.

[23] For example, in Ginzburg's (2012) *Dialogue Game Board* system, the assertion of *p* adds the polar question *p?* to the top of the Question Under Discussion (QUD) stack, under the assumption that asserting *p* is a proposal to add *p* to the common ground, which has to be accepted (often implicitly) by all interlocutors before moving on.

(53) [Context: after silently deliberating whether or not to order the house burger, I address the waiter who just came over.]

a. #Do you think I should do it?

b. #Does it come with fries?

c. #I'll have one medium-rare, please.

The non-linguistic context in this example does not serve up the intended referent with sufficient salience and as a result the use of *do it* anaphora, *it* or *one* is infelicitous. Furthermore, even within a context that enables the exophoric use of a certain referring expression, others may nonetheless be infelicitous. For example, in (54a), the son can use *that* felicitously to refer exophorically to his Dad's joke, whereas *it* is marked in the same context. Conversely, once the intended referent is mentioned explicitly by the phrase *this joke*, *that* is marked whereas *it* is now felicitous.[24]

(54) a. Dad: What do you call a dinosaur that's sleepy? A dino-snore.
Son: {#It | That}'s not funny!

b. Dad: When does a joke become a "dad joke?" When the punchline is a parent.
Son: I like this joke!
Dad: {It | #That}'s funny, right?

Obviously, it would be a mistake to argue based on this difference that *it* is not a form of discourse reference simply because it fails a within-context comparison with another referential expression. Rather, as discussed in Section 2.2.1, different referring expressions are felicitous under different conditions (see (48) for differences between *it* and *that*), and (54a-b) demonstrate their ability to refer exophorically is constrained by the extent to which the non-linguistic context satisfies

[24]I consider the reference to the joke exophoric because, even though the joke is *performed* linguistically, it is not *mentioned* linguistically. Consequently, it enters into the discourse model by virtue of the fact that both interlocutors are (mutually) aware of the joke and can thus refer to it exophorically. Alternatively, one could imagine a non-linguistically performed joke, for example in the form of a clown's performance, and the same pattern would emerge: *that*, but not *it*, is felicitous exophorically, and the reverse is true once the referent is denoted linguistically.

these conditions. In other words, Miller and Pullum's (2013) conclusions about exophoric VP-ellipsis appear to generalize to non-elliptical cases of exophora, and that generalization follows straightforwardly from referential theories of ellipsis.

To summarize, H&S argued that the possibility of exophoric reference is diagnostic of discourse reference and that exophoric ellipsis is infelicitous (with the exception of conventionalized utterances). However, it appears that both VP-ellipsis and sluicing *can* be used exophorically as long as the non-linguistic context provides sufficient support and meets their discourse conditions. Furthermore, the fact that exophora requires a higher degree of contextual support than cases in which the intended referent is introduced linguistically is not restricted to ellipsis: non-elliptical forms of discourse reference are also infelicitous when the non-linguistic context fails to raise the intended referent to a sufficient level of salience. Both of these parallels follow straightforwardly from analyzing ellipsis as a form of discourse reference. By contrast, the possibility of exophoric ellipsis presents a serious challenge for IDENTITY theories, which predict all cases of (non-conventionalized) exophora to be fully ungrammatical.

Multiple "split" antecedents

Discourse referential devices all have the ability to "pick up" entities introduced by multiple "split" antecedents. When they do, the meaning the referring expression acquires anaphorically reflects all of its antecedents in some way and cannot be reduced to a single antecedent. Consider the following examples:

(55) a. A: I've heard Susan might vote to allow witnesses at Donald's impeachment trial and Mitt said he will even vote to remove him from office.
B: I heard the same thing, but I doubt that either of <u>them</u> is actually going to <u>do it</u>.

b. Jack failed chemistry and Sara had to drop out of her arts class. Neither of <u>them</u> was surprised that <u>it</u> happened, but their parents were.

In (55a), *them* refers to the set of individuals introduced separately by the NPs *Susan* and *Mitt*, and the referent of *do it* is similarly interpreted as a combination of the activities Susan and Mitt were said to be considering, distributed via a "respectively" relation. Likewise, the pronoun *it* in (55b) appears to simultaneously refer to two events: Jack failing chemistry; and Sara dropping out of arts.

Whatever inferential mechanism enables these split-antecedent interpretations (see e.g., Nash-Webber, 1977; Hardt, 1992; Baker, 2007), the same mechanism appears to be at play in interpreting VP-ellipsis and sluicing: just like other referring expressions, ellipsis sites can felicitously acquire interpretations that depend on multiple antecedents:

(56) a. Mary wants to go to Spain and Fred wants to go to Peru, but because of limited resources only one of them can (go to the place she or he wants to go to, respectively).
(Webber, 1978, ch. 4, ex. 9)

b. Wendy is going to Spain and Bruce is going to Crete, but in neither case do I know why (Wendy is going to Spain and Bruce to Crete, respectively).
(Nash-Webber, 1977, ex. 1)

c. Rachel said she saw something, and her brother said he heard something, but neither of them could identify {it | what (they heard/saw, respectively)}.

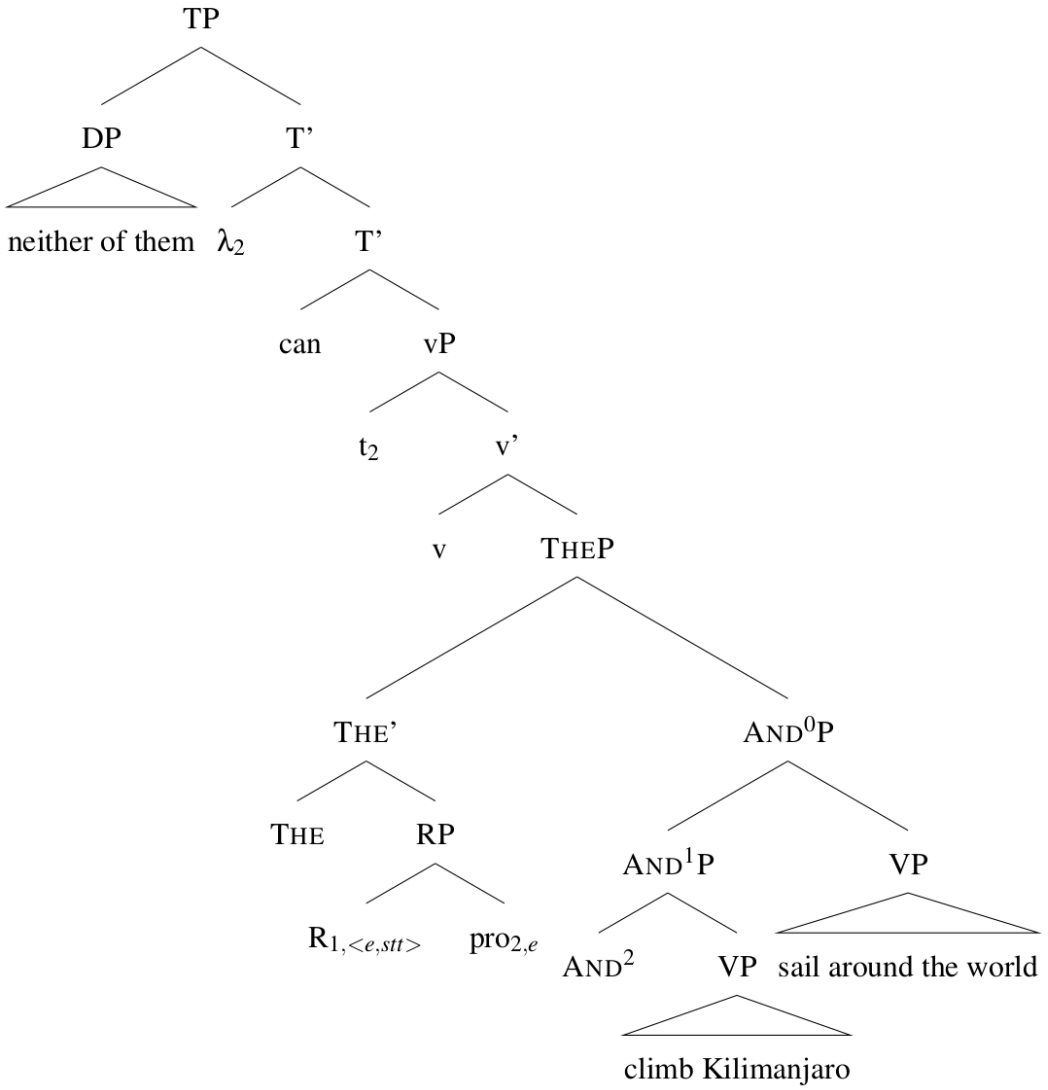
Just as the referring expressions in (55), the ellipsis sites in (56) depend for their interpretation on multiple antecedents: the VP-ellipsis in (56a) refers to an abstraction of the meanings introduced by the two antecedent VPs (both *Spain* and *Peru* are recognized as *places*), and the sluice in (56b) questions the reasons Wendy and Bruce have for going to Spain and Crete, respectively. The context in (56c) is suitable for split-antecedent reference both by the pronoun *it*, as well as the sluiced question *what they saw/heard, respectively*.

The ability of elliptical utterances to refer to discourse entities that are inferred by com-

bining entities introduced by separate linguistic antecedents is unsurprising on a referential theory of ellipsis. For IDENTITY theories, on the other hand, this fact raises a serious challenge: how can elided material be identical (at any level of representation) to multiple linguistic antecedents that are not identical to each other? Indeed, while discussion of split-antecedent ellipsis is largely absent from the literature, one influential attempt of explaining them from an IDENTITY perspective, due to Elbourne (2008), has to make numerous assumptions to make it work. In brief, Elbourne takes an approach that is similar to the "fine-tuning" approaches outlined in Section 2.1.2: he argues that the meaning of ellipsis clauses with multiple antecedents, like (57a), derives from an underlying syntactic structure like the one in (57b):

(57) a. Bob wants to sail around the world and Alice wants to climb Kilimanjaro, but neither of them can (sail around the world or climb Kilimanjaro, respectively).

b.



(Elbourne, 2008, ex. 45)

While everything below the auxiliary *can* is elided, Elbourne stipulates that only the nodes labeled *VP* are subject to the IDENTITY requirement, whereas the other structural elements, which are needed to derive the attested interpretations, can be elided "freely" even though they are not provided by any of the antecedents or any other part of the linguistic context. Furthermore, the semantics of AND^2, R_1, pro_2, and THE must be carefully defined so as to give rise to the "respectively reading" of the ellipsis clause.

A slightly different approach is due to Frazier and Duff (2019) who concede that split-

antecedent ellipsis (they focus specifically on VP-ellipsis) does violate the IDENTITY condition. In other words, they reject the notion of re-defining the IDENTITY relation in an effort to account for split-antecedent cases, and they further argue, contra Frazier (2013), that they do not exhibit the behavioral profile of syntactic repair, either (Arregui et al., 2006; Frazier, 2013). Instead, they propose that comprehenders' fleeting memory for past syntactic material (see e.g., Futrell, Gibson, & Levy, 2020) makes it impossible to retrieve the exact syntactic form of relatively distant antecedents. In contexts with multiple antecedents, ellipsis is thus exempt from IDENTITY simply because the syntactic form of the more distant antecedent is unavailable due to memory constraints. In such cases, the ellipsis clause instead derives its meaning through an inferential process Frazier and Duff term "accommodation," although explaining the mechanism behind accommodation is left as an objective for future research.

Both of these approaches are problematic for two reasons: first, they are based purely on the data they seek to explain and are thus not independently motivated. Specifically, Elbourne identifies the inferential gap between the individual antecedents and the interpretation of the ellipsis site and then designs a syntactic computation that derives this interpretation. Similarly, Frazier and Duff propose that all and only cases with multiple antecedents be excepted from the IDENTITY requirement, thus likewise tailoring their solution to the data in question. Secondly, fashioning an ellipsis-specific explanation misses the generalization that split-antecedent interpretations are not unique to ellipsis and leaves open the question how pronouns and other non-elliptical forms of reference acquire the same kinds of interpretations. Referential theories, on the other hand, naturally derive this generalization and do not need to make any special-purpose representational assumptions to explain the possibility of split-antecedent ellipsis.

Non-local antecedents

Another characteristic of discourse reference is that referring expressions and their antecedents may, in principle, be several sentences apart. As Sag and Hankamer (1984) point

out, this fact supports the notion that discourse referents are represented separately from their linguistic antecedents (in the interlocutors' discourse model) because comprehenders' memory for past linguistic material is fleeting (e.g., Jarvella, 1971; Gibson & Thomas, 1999; Futrell & Levy, 2017; Futrell et al., 2020): once a discourse entity has been introduced into the discourse model, it can be accessed even if the antecedent itself cannot reliably be retrieved from memory. Consider the examples of non-elliptical referring expressions in (58), and of sluicing and VP-ellipsis in (59).

(58) a. Bob hid the candy from his parents. They were strict and tried to make sure he didn't eat too many sweet things. But this time, he had hidden it well and there was no way they would find it.

b. Even though the suspect confessed to killing the victim, it wasn't a straightforward confession. She said she feared for her life and that she was only defending herself. If that's true and she didn't do it in cold blood, she might not end up being prosecuted for murder.

(59) a. Cindy didn't want to do her homework. She was tired and not in the mood. Besides, math was her least favorite subject. She knew she had to (do her homework) eventually, but she was determined to delay it as much as possible.

b. A: Who visited your uncle in the hospital?
B: I did.
A: When was that?
B: Last Thursday.
A: And who else (visited your uncle)?

These examples contrast with the following comparison from Sag and Hankamer (1984; adapted from Grosz, 1977), which they consider evidence that ellipsis does *not* involve discourse reference:

(60) E: Good morning. I would like for you to reassemble the compressor... I suggest you begin by attaching the pump to the platform...(other tasks).

A: All right. I assume the hold in the housing cover opens to the pump pulley rather than to the motor pulley.

E: Yes, that is correct. The pump pulley also acts as a fan to cool the pump.

A: Fine. Thank you. All right, the belt housing cover is on and tightened down. (30 minutes and 60 utterances after beginning.)

a. E: Fine, I knew you would be able to do it. [meaning: reassemble the compressor]

b. E: Fine, I knew you would be able to #(reassemble the compressor).

c. E: Fine. Now you know how #(to reassemble the compressor).

In this example, *do it* has no trouble referring to the event originally introduced 30 minutes earlier, but both VP-ellipsis and sluicing are infelicitous in the same context. However, while Sag and Hankamer (1984) interpret this within-context comparison as indicating that ellipsis with non-local antecedents is always impossible, we have seen in the context of exophora above that individual within-context comparisons may be misleading. Indeed, Hardt (1990) found that approximately 5% of cases of VP-ellipsis in the Brown corpus featured antecedents going at least two sentences back, and similarly Rønning, Hardt, and Søgaard (2018) report that about 1% of sluices in Anand and McCloskey's (2015) corpus have antecedents going three or more sentences back. Consistent with this, examples like those in (59) suggest that both VP-ellipsis and sluicing *can* refer to entities introduced by non-local antecedents, even though there is no guarantee that they can do so in the same contexts as other referring expressions.

Cataphora

It is well-known that pronouns, as well as VP-ellipsis and sluicing, can be used cataphorically, i.e. in contexts in which the referring expression precedes the antecedent, which is then

referred to as the "catacedent" and highlighted throughout this section in *italics*.

(61) a. If he wins re-election in November, *Donald* may avoid criminal prosecution because the statute of limitations will have run out by the time he leaves office.

b. If you promise to do it tomorrow, I'll *cook dinner today*.

c. Even though it happened too quickly for anyone else to react, the Jedi was able to parry *the attack*.

(62) a. And I know that as much as some of you might want me to, it's 2018 and I'm a woman so you cannot *shut me up*—unless you have Michael Cohen wire me $100,000.[25]

b. Until today, [Joe Biden] had not even campaigned in 1 of the 15 Super Tuesday states in over a month. Like Hillary did, he just *assumed he'd be crowned King everywhere*. Hardly has offices or staff anywhere.[26]

c. He didn't know why, for instance, but *when he flew at altitudes less than half his wingspan above the water, he could stay in the air longer, with less effort*.

(Hinds & Okada, 1975, citing Richard Bach's *Jonathan Livingston Seagull*)

This parallel between ellipsis and non-elliptical referring expressions extends further to the conditions under which cataphora is felicitous. Specifically, it requires the referring expression to be embedded in a subordinate clause, as shown for both entity-level pronouns and VP-ellipsis in the following examples due to Kehler (2019):

(63) a. If he makes a statement criticizing President Putin, *Obama* will make a fool of

[25]This example is from Michelle Wolf's speech at the 2018 White House Correspondents Dinner, available at https://youtu.be/L8IYPnnsYJw?t=2m26s at the time of writing. Besides the use of cataphora, this example is interesting because it features an argument-structure mismatch between inchoative and causative uses of *shut up*, which should render it ungrammatical according to IDENTITY theories (Chung, 2006; Chung et al., 2011; Lipták, 2015).

[26]Tweet from Shaun King from March 1, 2020; available at https://twitter.com/shaunking/status/1234178868677771264?s=09 at the time of writing.

himself.

b. #He will make a fool of himself, if *Obama* makes a statement criticizing President Putin.

(64) a. If McCain will (make a statement criticizing President Putin), Obama will *make a statement criticizing President Putin*.

b. Obama will #(make a statement criticizing President Putin), if McCain will *make a statement criticizing President Putin*.

The key intuition is that referring expressions are effortlessly interpreted as co-referential with their catacedent (in *italics*) only when they are embedded in a subordinate clause, as in (63a) and (64a). By contrast, when the referring expression is not in a subordinating environment, comprehenders may find themselves looking for another antecedent or perhaps even a situationally evoked referent. This dispreference for cataphoric co-reference in non-subordinating environments is the diagnostic that is at stake here.

(63) and (64) establish that VP-ellipsis patterns with entity-level pronouns in this regard, and the same appears to be true of sluicing:

(65) a. Even though he remembered when, John forgot where *he was supposed to meet Bill*.

b. John forgot when #(he was supposed to meet him), and he also forgot where *he was supposed to meet Bill*.

IDENTITY theories of ellipsis are consistent with the possibility of cataphoric ellipsis because the IDENTITY condition does not care *where* an identical antecedent is found, only *that* one be available. However, the fact that both elliptical and non-elliptical cases of cataphora are subject to the same subordination constraint remains unexplained under that view. It follows straightforwardly, on the other hand, from referential theories: if the two are governed by the

same mechanism, it is unsurprising that they would obey the same constraints on co-reference establishment. Indeed, Frazier (2013, p. 497) concedes that the shared subordination requirement is "[p]erhaps the most persuasive of the arguments analogi[z]ing ellipsis to anaphora."

Triggering "sloppy" readings

Another well-known fact about elliptical utterances is that they can be ambiguous between two interpretations known as "strict" and "sloppy" readings (Ross, 1969; Sag, 1976; Williams, 1977; Dalrymple et al., 1991; Hardt, 1993; Fox, 1999; Kehler, 2016), as illustrated in the following examples.

(66) a. Dan called his sister and Bill did (call his, i.e. Bill's, sister), too.

b. 5 is equal to itself and 7 is (equal to itself, i.e. 7), too.

(adapted from Rooth, 1992, ex. 5)

In both cases, the underlined expression in the antecedent clause is free to refer to a different entity in the ellipsis clause, which Ross (1969) termed "sloppy identity." While sloppy interpretations are by no means unique to ellipsis,[27] they are available if, and only if, the antecedent contains a discourse-referential device, such as *his* and *itself* in (66): if these pro-forms are replaced with full NPs that do not engage the referential system in the same way, the sloppy interpretation is no longer available, or at least much less so (Dalrymple et al., 1991):

(67) a. Dan called Dan's sister and Bill did #(call his, i.e. Bill's, sister), too.

b. 5 is equal to 5 and 7 is #(equal to itself), too.

Even though the antecedent clause has the same truth conditions as before, the absence of a

[27] See Tancredi (1992) and Kehler (1993a) for examples of "sloppy" interpretations under deaccenting, *do it* anaphora, as well as entity-level pronouns known as "paycheck pronouns" or "pronouns of laziness" (Geach, 1962; Karttunen, 1969; Hardt, 1994).

referring expression in the antecedent of the ellipsis site prevents the sloppy interpretation that was previously available. Generalizing this contrast, we can thus leverage the (un)availability of a sloppy interpretation for some linguistic expression, say β, to determine whether some other expression—let's call it α—involves discourse reference by embedding it in the antecedent of β, following the diagnostic in (68b):

(68) a. $\ldots[\ldots\alpha\ldots]_i\ldots\beta_i\ldots$

b. If α is contained in the antecedent of β, as shown in (68a), then β has a sloppy interpretation with respect to α if, and only if, α is a form of discourse reference.

When applied to (66) and (67), this diagnostic correctly concludes that *his* and *itself* are discourse-referential, whereas *Dan* and *5* are not. Crucially for our purposes here, however, we can also apply it to cases of VP-ellipsis and sluicing (by replacing α with an ellipsis site), as well as their unelided counterparts, to determine whether they engage the system of discourse reference in the same way that non-elliptical pro-forms do. Consider the following example, versions of which were first discussed by Hardt (1994) and later re-discovered by Schwarz (2000).

(69) a. When Harry drinks, I always conceal my belief that he shouldn't (drink). But when he gambles, I often can't (conceal my belief that he shouldn't gamble).

b. When Harry drinks, I always conceal my belief that he shouldn't drink. But when he gambles, I often can't #(conceal my belief that he shouldn't gamble).

The meaning of the VP-ellipsis site in (69a) changes between the antecedent clause ("...he shouldn't drink") and the ellipsis clause ("...he shouldn't gamble"), giving rise to a "sloppy" interpretation. This suggests that VP-ellipsis engages the discourse reference system in a way that its unelided counterpart with the same meaning, as in (69b), does not.[28]

[28]The sloppy reading re-emerges if the ellipsis site is replaced with *do it*, which further underscores the parallel between VP-ellipsis and non-elliptical referring expressions.

An analogous example can be constructed for sluicing, as shown in (70). As before, the sluicing variant in (70a) makes the sloppy reading *explain why he likes <u>cake</u>* available, but it disappears when the sluice is replaced with an overt variant of the same question, as in (70b). Finally, (70c) demonstrates that the pronoun *it* patterns with the sluice: it, too, allows for a sloppy reading of the subsequent VP-ellipsis.

(70) a. Susan likes steak and Bill prefers cake. Susan can't explain why (she likes steak) and Bill can't (explain why he likes cake), either.

b. Susan likes steak and Bill prefers cake. Susan can't explain why she likes steak and Bill can't #(explain why he likes cake), either.

c. Susan likes steak and Bill prefers cake. Susan can't explain it and Bill can't (explain why he prefers cake), either.

Since *why* sluices behave differently from other sluices in a variety of ways (e.g., Merchant, 2001), it is worth emphasizing that the above pattern is not limited to *why* questions:

(71) Billy wants a new bike for Christmas and his sister wants a new skateboard.

a. Billy knows exactly what type (of bike he wants), and his sister does (know exactly what type of skateboard she wants), too.

b. Billy knows exactly what type of bike he wants, and his sister does #(know exactly what type of skateboard she wants), too.

(72) Donald announced that he will hold a rally in May and Melania said that she will host a charity dinner in June.

a. Donald didn't specify where (he will hold the rally), and Melania didn't (specify where she will host the charity dinner) either.

b. Donald didn't specify where he will hold the rally, and Melania didn't #(specify

where she will host the charity dinner) either.

Both VP-ellipsis and sluicing, then, appear to pattern with other pro-forms in their ability to trigger sloppy interpretations of "downstream" anaphoric elements. Once again, this similarity is unsurprising if elliptical utterances and non-elliptical forms of discourse reference are governed by the same underlying mechanism. On the other hand, this behavior is problematic for IDENTITY theories: since elliptical utterances and their unelided counterparts are expected to be identical in all respects except phonologically, it is surprising that VP-ellipsis and sluicing sites trigger sloppy readings if unelided VPs or clauses in the exact same contexts do not (Hardt, 1994, and many others following him). Acknowledging this issue, Tomioka (2008) proposes an exception to the IDENTITY condition with respect to material in embedded ellipsis sites by stipulating that the mechanism that checks IDENTITY "skips" any material in such positions. A similar proposal has been put forward by Merchant (2004) who argues that the elided material in the embedded VP-ellipsis site in cases like (69) must be *do that*, which then receives a sloppy interpretation "downstream" just like other pro-forms do. However, those analyses are problematic for obvious reasons: First, the ellipsis of *do that* in Merchant's example itself violates IDENTITY, at least at a lexico-syntactic level. Secondly, extending this analysis to sluicing would require further stipulating the IDENTITY-free ellipsis of other material and it is not clear to me what material that would have to be since there is no *do that* equivalent in English for anaphorically recovering the meaning of clauses embedded under interrogative *wh*-phrases (except for sluicing itself, of course). Third, any IDENTITY theory of VP-ellipsis that permits "do that" to be elided IDENTITY-free will overgenerate in many other cases, including unacceptable cases of passive-active VP-ellipsis. Fourth, both Tomioka's and Merchant's proposals are entirely post-hoc, tailored specifically to the observations they purport to explain, which undermines their explanatory value. Finally, being designed specifically for ellipsis, both proposals miss the generalization that the ability to trigger sloppy readings is not limited to ellipsis, leaving the same behavior unexplained with respect to

pronouns and other referential expressions.

Inferred referents

Recall from Section 2.2.1 that entities can be entered into the discourse model inferentially. The following examples, repeated from (44), illustrate this point.

(73) a. Susan called Becky to discuss their dinner plans.

b. Harry threw up and Sam stepped in it.

c. When Jack was kidnapped, they kept him in a dark room for days.

While the referents in such cases are not denoted directly by any particular part of the linguistic context, they nonetheless exhibit a certain "closeness" to it. For example, *they* in (73a) cannot be interpreted as referring to any arbitrary set of individuals; instead, its meaning is restricted to the set composed of its ("split") antecedents *Susan* and *Becky*. Likewise, *it* and *they* in (73b-c) are interpreted in a way that is closely related to the puking and kidnapping events, respectively, which are introduced explicitly. Furthermore, the inferences that introduce entities into the discourse model without denoting them explicitly must be supported by (and are thus constrained by) the interlocutors' shared world knowledge. (73b), for example, requires knowledge about the product of a puking event, and the referent in (73c) is inferred by recognizing that kidnapping events necessarily involve agents.[29]

This provides us with a final diagnostic property, which also helps set the stage for the discussion in the next section: if VP-ellipsis and sluicing are forms of discourse reference, we

[29] In that respect, these inferences resemble bridging inferences (e.g., Clark, 1975), which serve to accommodate a definite NP referring to a discourse-new entity based on its relation to a discourse-old entity, as supported by interlocutors' shared world knowledge. For example, in (ia) *the murderer* is interpreted specifically as the person who murdered John, and *the knife* in (ib) as the weapon used in the particular event denoted by the antecedent clause.

(i) a. John died yesterday. The murderer got away. (Clark, 1975, ex. 23)

b. John was murdered yesterday. The knife lay nearby. (Clark, 1975, ex. 24)

should expect (i) that inferential ellipsis resolution is, in principle, possible, i.e. that there are cases in which the meaning of the ellipsis site goes beyond the meaning that is denoted by linguistic antecedent; and (ii) that the inferences in question are constrained by world knowledge just as they are in the case of non-elliptical reference. Both of those predictions seem to be borne out. First, as discussed above, the interpretation of ellipsis clauses with split antecedents, repeated in (74), exhibits a similar "closeness" in meaning to its antecedents as plural entity-level pronouns. Furthermore, the inferences that establish split-antecedent interpretations often involve an abstraction in line with the interlocutors' world knowledge, such as the knowledge that *Spain* and *Peru* are both places, or that *wanting to go somewhere* and *planning to do so* often come together.

(74) Mary wants to go to Spain and Fred wants to go to Peru, but because of limited resources only one of them can (go to the place he or she is planning to go to).

(Webber, 1978, ch. 4, ex. 9)

Secondly, various other cases of "mismatch" between the meaning of the ellipsis site and its antecedent (see Section 2.1.1) appear to be facilitated by world knowledge as well. For example, the following lexical mismatches from Kehler (2002a) clearly require the knowledge that *boyfriend* and *girlfriend* (and *parent* and *child*) can be understood as complementary relations, as well as other assumptions, such as that multiple "schoolboys" are unlikely to have the same girlfriend (hence the plural *their girlfriends*), and so forth.

(75) a. Mary's boyfriend gave her his school picture, just as all schoolboys do (give their girlfriends their school picture).

b. A: Bob's mother cleans up after him all the time.
B: I'm surprised; most parents these days won't (clean up after their children).

Similarly, the inference in (76a) requires the recognition that Mary's role in the event of *Irv and Mary dancing together* would be to *dance with Irv*, and the sluiced question in (76b) is interpreted in accordance with the knowledge that a *when* question is relevant in the context of agreeing to have coffee.

(76) a. Irv and Mary want to dance together, but Mary can't (dance with Irv), since her husband is here. (Webber, 1978, ch. 4, ex. 8, parentheses added)

b. Coffee sounds good. When (should we have coffee)?

(adapted from Ginzburg, 1992, ex. 303a)

In all of these cases, the interpretation of the ellipsis clause is "anchored" to the linguistic antecedent (or some other part of the linguistic context), and the inferences that take it beyond the meaning of the antecedent are constrained and supported by the interlocutors' mutually held conceptual knowledge. Inferential ellipsis resolution, its relation to acceptability, and the role of world knowledge will be discussed in more detail in Chapters 4 and 6. For present purposes, the key observation is that VP-ellipsis and sluicing sites can be interpreted inferentially and that inferential ellipsis resolution appears to be constrained in ways similar to non-elliptical reference to inferred discourse entities.

Beyond VP-ellipsis and sluicing

In this section, I considered six diagnostic properties of discourse reference: exophora, split-antecedent reference, non-local antecedents, cataphora, sloppy interpretations, and inferred referents. The fact that VP-ellipsis and sluicing exhibit all of those properties provides strong support for the claim that they are governed by the same underlying mechanism as other forms of discourse reference. Furthermore, several of these properties raise serious challenges for IDENTITY theories of ellipsis.

Table 2.2: Summary of anaphoric properties across discourse referential devices. (✗) indicates that the diagnostic in question is not fully applicable.

Phenomenon	VPE	Sluicing	NPE	NCA	Gapping
Exophora	✓	✓	✓	✓	✗
Split antecedents	✓	✓	✓	✓	✗
Non-local antecedents	✓	✓	✓	✓	✗
Cataphora	✓	✓	✓	✓	(✗)
"Sloppy" readings	✓	✓	✓	✓	(✗)
Inferred referents	✓	✓	✓	✓	✗

In keeping with the overall theme of this thesis, the discussion above has focused on VP-ellipsis and sluicing, but it is worth emphasizing that several other forms of ellipsis pass these diagnostics as well. In what follows, I provide examples for NP-ellipsis, Null Complement Anaphora (NCA), and Gapping; as shown in Table 2.2, NP-ellipsis and NCA do seem to engage the referential system.[30] Gapping, on the other hand, does not exhibit any of the diagnostic properties and thus serves as an informative baseline for what non-referential ellipsis may look like.

(77) NP-ellipsis

a. Exophora: [Context: In a parking lot.] Where's your brother's (car)? [31]

b. Split antecedents: John needs a hammer. Mary needs a mallet. They're going to borrow Bill's (hammer/mallet, respectively). (Elbourne, 2008, ex. 19)

c. Non-local antecedents: Billy has been thinking about cake all week. It was a busy week and he didn't have time to go grocery shopping. Since he's going today, however, he is hopeful that he will finally be able to have some (cake).

d. Cataphora:

[30] Note that this conclusion is consistent with Hankamer and Sag's classification of NCA as a "model-interpretive" ("deep") anaphor. NP-ellipsis, on the other hand, is often analyzed as a surface anaphor that is subject to IDENTITY (e.g., Elbourne, 2001, 2008; Merchant, 2019), which leaves these anaphoric properties of NP-ellipsis unexplained.

[31] Khullar, Majmundar, and Shrivastava (2020) suggest that exophoric NP-ellipsis may be very common indeed: they find 946 cases of NP-ellipsis in the *Cornell Movie Dialog* dataset, 508 of which they classified as exophoric (54%).

(i) Even though she accepted Bob's (apology), Jessie didn't accept Bill's apology.

(ii) Jessie didn't accept Bill's #(apology), and I don't think she'll accept Bob's apology.

e. Triggering sloppy interpretations: Susan wants cake and Bill would love some steak. Susan isn't going to have any (cake) and Bill won't (have any steak), either.

f. Inferred referents: When the kids all threw up, I'm the one that had to clean Johnny's (vomit) up.[32]

(78) Null Complement Anaphora (NCA)

a. Exophora: [Context: Teenager comes home after curfew.] Parent: I do not approve (of your coming home so late).

b. Split antecedents: Billy left through the window and Gracie snuck out through the garage. Because they were quiet, their parents didn't notice (that they left through the window/garage, respectively).

c. Non-local antecedents:

A: Do you know the final score of the game today?

B: Which game?

A: The Champions League semi-final, of course.

B: The Liverpool game?

A: Yes.

B: I don't know (the final score of that game), sorry.

d. Cataphora:

(i) Even though she tried (to set up the zoom meeting), Karen didn't manage to set up the zoom meeting.

(ii) Karen tried ?(to set up the zoom meeting), and eventually she managed to set up the zoom meeting.

[32]I am grateful to Andy Kehler for providing this example.

e. Triggering sloppy interpretations: Donald was going to play golf and Melania was going to go shopping. Donald's Secret Service detail didn't approve (of his plan to play golf) and Melania's didn't (approve of her plan to go shopping), either.

f. Inferred referents: One hostage didn't know the answer to the kidnappers' question and the other simply refused (to answer the question).

(79) Gapping

a. Exophora: [Context: right after a car runs a red light.] One pedestrian to another: Yesterday, a TRUCK #(ran) a red light.

b. Split antecedents: Leslie saw the first car coming and Beto heard it. Neither of them #(saw/heard, respectively) the second one.

c. Non-local antecedents: Nina called her father on Monday. On Tuesday she was busy all day. That's why her sister #(called) her mother.

d. Cataphora:[33]

(i) Even though Susan #(accepted) Bob's apology, Jessie didn't accept Bill's.

(ii) Susan #(accepted) Bob's apology, and Jessie accepted Bill's.

e. Triggering sloppy interpretations: The women all called their friends and the men texted theirs. Specifically, Susan said that she called her friends and her friends (called) theirs, and Jack did #(say that he texted his friends and his friends texted theirs), too.[34]

f. Inferred referents: Irv and Mary want to dance together, and Jack #(wants to dance with) Sue.

The data in (79) highlight a series of disanalogies between Gapping on one side and sluicing, VP-ellipsis, NCA, NP-ellipsis, and non-elliptical referential devices on the other. This

[33]The cataphora diagnostic requires subordination, which is independently prohibited for Gapping.

[34]The relevant examples are impossible to construct because Gapping sites (i) cannot be embedded and (ii) must be immediately preceded by their antecedents.

picture therefore casts doubt over theories of ellipsis that offer unified explanations of Gapping and other forms of ellipsis by modeling them as phenomena governed by the same underlying mechanism (e.g., Culicover & Jackendoff, 2005, 2012; Goldberg & Perek, 2019): if, as the observations here suggest, they are enabled by architecturally distinct mechanisms, then it may be a mistake to try to explain them through the same theoretical constructs.

2.2.3 Inferential reference resolution and the morphosyntactic form of the antecedent

Section 2.2.1 described various ways in which linguistic antecedents affect discourse reference. First, the semantic object it denotes is introduced into the discourse model and becomes available for subsequent reference. Secondly, its morphosyntactic properties constrain what expressions can be used to refer to the entity it introduces, leading to agreement in the morphosyntactic marking of, for example, gender and number. Third, as discussed at the end of Section 2.2.2, the linguistic antecedent serves to "anchor" inferential reference resolution, allowing interlocutors to identify and refer to discourse-new entities through their relation with the antecedent. This section expands on the latter point by showing that, contra H&S (Hankamer & Sag, 1976; Sag & Hankamer, 1984), the felicitous use of discourse-referential expressions *does* depend on the morphosyntactic properties of the antecedent in various ways, especially when referring to inferred referents.[35] In the context of these observations about non-elliptical forms of reference, I will then argue that mismatches between VP-ellipsis and sluicing sites and their linguistic antecedents can be analyzed analogously as a function of the accessibility of the intended referent, which in turn is affected by the morphosyntactic form of the antecedent.

[35] In fact, this conclusion follows from the core assumption that discourse reference is grounded in a "contract" between interlocutors that requires that referents be in common ground (Nash-Webber, 1977, *Abstract*, p. 0): "This contract requires that if the speaker uses an anaphoric expression whose [...] referent was inferentially derived, the listener both can and will make the same inference. Insofar as it is shown that many of these inferences rely on one of the few things explicitly available to both speaker and listener alike - i.e., the form of the utterance - the identification of a sentence's formal properties become a matter of cognitive concern."

Consider first the following cases of infelicitous discourse reference, taken from Ward et al. (1991):

(80) a. #Max is an orphan and he deeply misses them [= his parents]. (Postal, 1969, ex. 3a)

b. Fritz is a cowboy. # He says they [= the cows] can be difficult to look after. (Ward et al., 1991, ex. 23a)

c. Dom's clothes are absolutely elephantine. # Indeed you could almost lose one [= elephant] in them. (Ward et al., 1991, ex. 23d)

Early analyses of such cases maintained that complex NPs are 'anaphoric islands' and impose a categorical syntactic constraint that prevents any elements embedded in them from participating in anaphoric dependencies (Postal, 1969). By contrast, Ward et al. (1991) argued that reference into and out of complex NPs depends on the (gradient) accessibility of the intended referent. For example, cases like the following are perfectly felicitous:

(81) a. Do parental reactions affect their [= the parents'] children? (Ward et al., 1991, p. 469)

b. Although casual cocaine use is down, the number of people using it [= cocaine] routinely has increased. (Ward et al., 1991, ex. 22a, originally heard on the news)

According to Ward et al. (1991), accessibility depends on the degree to which computing the meaning of the complex NP requires comprehenders to access the meaning of the intended referent. For example, whereas *orphan*, *cowboy* and *elephantine* in (80) have conventionalized meanings, NPs like *parental reactions* and *cocaine use* in (81) are interpreted compositionally and thus require comprehenders to access the meanings of *parent* and *cocaine*, thereby making them accessible for subsequent reference.

Referent accessibility further depends on the morphological transparency of the antecedent with respect to the intended referent:[36] while *France* is readily accessible for reference via *there* in (82a-b), *Denmark* and *the Netherlands* appear to be less so in (82b-c) (Kehler, personal communication).

(82) a. Jean is from France, but he hasn't been there [= in France] in years.

b. ?Jean is French, but he hasn't been there [= in France] in years.

c. ??Jean is Danish, but he hasn't been there [= in Denmark] in years.

d. ??Jean is Dutch, but he hasn't been there [= in the Netherlands] in years.

While much research on reference has focused on entity-level reference, similar accessibility facts can be observed for referential expressions targeting events. For example, Ward and Kehler (2005) and Kehler and Ward (2007) show that *do so* can felicitously refer to events introduced by nominal antecedents, but only to the extent that the events in question are sufficiently accessible:

(83) a. One study suggests that almost half of young female smokers do so in order to lose weight. (Kehler & Ward, 2007, ex. 18)

b. The greatest teachers do so by example. (Kehler & Ward, 2007, ex. 38)

c. #Most professors will do so for hours even when no one is listening. (Kehler & Ward, 2007, ex. 22)

d. #In my opinion, our governor does so better than the last one did. (Kehler & Ward, 2007, ex. 23)

[36]For simplicity, I am glossing over the fact that Ward et al. (1991) further distinguish morphological transparency from "lexical relatedness" to account for the anaphoric accessibility of *two* based on the mention of *second* in the following example:

(i) This is the second time in as many [= two] weeks. (Ward et al., 1991, ex. 10)

The idea that reference resolution depends on the (gradient) accessibility of the referent is consistent with the facts around exophora and inferred referents we have seen above. In order for a referring expression to be felicitous, both interlocutors must be able to recognize the intended referent as mutually known, which is straightforward when it is denoted explicitly by an antecedent that both the speaker and the listener are (mutually) aware of. By contrast, entities that are situationally evoked or inferred (including ones that are introduced from within 'anaphoric islands') are accessible only to the extent that the non-linguistic context and interlocutors' shared world knowledge compensate for the lack of an antecedent. As a result, inferential and exophoric reference resolution is expected to be variably acceptable, and, as we will see next, this kind of gradience conditioned by accessibility is not restricted to non-elliptical reference but can also be seen in cases of VP-ellipsis and sluicing with "mismatching" antecedents, i.e. cases in which the meaning of the ellipsis site is not reducible to the meaning introduced by the antecedent.

One well-studied type of mismatch case is VP-ellipsis with nominal antecedents (Hardt, 1993; Johnson, 2001; Miller & Hemforth, 2014, among others), which were briefly discussed in the context of IDENTITY theories in Section 2.1.2. While the mismatch renders such cases categorically ungrammatical according to IDENTITY theories, referential theories predict that they should be variably acceptable depending on the accessibility of the intended referent. Consistent with this prediction, Miller and Hemforth (2014) show based on corpus data and experimental data that cases of nominal-antecedent VP-ellipsis are acceptable only to the extent that the antecedent NP raises a "concealed question," as in the following examples (concealed questions in [brackets] added by me).

(84) a. Mubarak's survival [= whether he will survive] is impossible to predict and, even if he does (survive), his plan to make his son his heir apparent is now in serious jeopardy. (Miller & Hemforth, 2014, ex. 1)

b. The integrity of the Senate depends on her participation [= whether she participates].

If she does (participate), . . . (Miller & Hemforth, 2014, ex. 10a)

c. The release of this information on the user's part depends on his consent [= whether he consents]. If he does (consent), . . . (Miller & Hemforth, 2014, ex. 10c)

In each of these examples, processing the antecedent NP requires comprehenders to access the meaning of a concealed polar question, making it thereby accessible for subsequent VP-ellipsis in exactly the same way that 'anaphoric islands' enable non-elliptical reference (when they do) according to Ward et al. (1991): in both cases, the referent is introduced as a by-product of processing the antecedent NP.[37] Furthermore, the effect of morphological transparency we observed in (82)—whereby *France* was more accessible from *French* than *the Netherlands* from *Dutch*—can also be observed in nominal-antecedent VP-ellipsis. Consider the following minimal pair from Merchant (2013a).

(85) a. That man is a robber, and when he does ?(rob places), he tries not to make any noise.

b. That man is a thief, and when he does #(steal things), he tries not to make any noise.

As discussed in Section 2.1.2, attempts at explaining this contrast (as well as the fact that nominal-antecedent VP-ellipsis is possible at all) under IDENTITY theories of ellipsis typically involve

[37] According to Miller and Hemforth's (2014) analysis, nominal-antecedent VP-ellipsis further requires that the information structure of the concealed question raised by the antecedent NP be the same as the information structure of the subsequent ellipsis clause. For example, while auxiliary-focus VP-ellipsis is felicitous in the context of a *polar* question, it is marked when the antecedent NP raises an alternative (wh-) question:

(i) That depends on her answer [= what her answer is]. If she does #(answer), . . .
(Miller & Hemforth, 2014, ex. 12a)

This is consistent with Kertz's (2013) analysis of VP-ellipsis with voice-mismatched antecedents as well as Miller and Pullum's (2013) analysis of exophoric VP-ellipsis (where there is *no* linguistic antecedent). All of these cases of VP-ellipsis in the absence of a suitable antecedent VP can thus be understood in terms of the accessibility of a suitable Question Under Discussion (Roberts, 1998; Ginzburg & Sag, 2000; Roberts, 2012). While the exact relation between a QUD and the accessibility of the VP meaning it contains remains a subject for future research, this approach is promising from the perspective of referential theories of ellipsis.

assuming that deverbal nouns underlyingly contain the VP they derive from and that that VP is available to serve as the antecedent for subsequent VP-ellipsis (Fu et al., 2001; Johnson, 2001; Merchant, 2013a). However, since this explanation is tailored specifically to VP-ellipsis, it cannot explain the fact that nominal-antecedent sluicing is possible as well and exhibits a similar degree of gradience:[38]

(86) a. Regarding Trump's impeachment, the only question is when (he will be impeached).
b. Regarding Trump's impeachment, the only question is why #(he will be impeached).
c. Regarding Trump's impeachment, the only question is who #(will impeach him).

The accessibility-based explanation of these facts, on the other hand, follows straightforwardly from the behavior of non-elliptical forms of discourse reference if we assume that both VP-ellipsis and sluicing engage the same underlying mechanism.[39]

While referent accessibility undoubtedly affects the felicity of discourse reference, it is unlikely to be the full story. One of the most convincing demonstrations of this fact comes from the following type of example, usually attributed to Barbara Partee:

(87) a. I dropped ten marbles and I found all but one of them. It must be under the sofa.
b. I dropped ten marbles and I found only nine of them. # It must be under the sofa.

In (87a), the use of *it* to refer to the tenth marble is perfectly felicitous: its referent is denoted by the antecedent NP *one of them*. In (87b), on the other hand, it is not denoted by any part of

[38] See Chapter 6 for a series of experiments on nominal-antecedent sluicing along with a more in-depth discussion of its implications for IDENTITY theories.

[39] Even more generally, morphological transparency further appears to affect non-referential linguistic expressions that depend on the discourse-Givenness of a certain meaning. For example, as pointed out to me by Andy Kehler (p.c.), VP preposing is only felicitous when the meaning of the VP is Given in the discourse, and the NP *rainfall* appears to satisfy this condition, whereas the near-synonymous NP *precipitation* does not:

(i) a. The weather forecast predicted heavy rainfall, and rain it did.
b. #The weather forecast predicted heavy precipitation, and rain it did.

the linguistic context and must instead be inferred, and as a result the use of *it* is infelicitous. Importantly, however, several aspects of the context conspire to make it maximally accessible, suggesting that referent accessibility *per se* is not enough for felicitous reference in this case: not only does the context plausibly raise the question *Did you find the marbles you dropped?*, the focus construction *only nine of them* specifically shifts attention to the last missing marble. As a result, if inferential reference resolution were reducible to the accessibility of the intended referent, we should expect the use of *it* to be perfectly felicitous in this context, but it clearly is not.

Since many of the theoretical constructs that have been found in previous research to constrain the felicitous use of discourse reference[40] are themselves poorly understood, it is important to develop independent operational definitions of them in order to avoid circularity in testing the predictions of referential theories of ellipsis. For example, Miller and Hemforth (2014) operationalized the extent to which nominal antecedents raise a concealed question through an experimental task that did not itself involve ellipsis. They presented participants with the antecedent clause (e.g., *Mubarak's survival is impossible to predict.*) along with a set of paraphrases, including a "polar paraphrase" that corresponds to the relevant concealed-question interpretation (e.g., *Whether or not Mubarak will survive is impossible to predict.*). Participants then rated each paraphrase in terms of how closely they matched the meaning of the antecedent clause of each item, and the average rating of the polar paraphrase was then used as to predict the acceptability of VP-ellipsis for each item.

2.3 Comparing referential and IDENTITY theories of ellipsis

In the previous sections, I introduced two theoretical approaches to ellipsis: IDENTITY theories, which maintain that ellipsis is governed by a special-purpose mechanism that enforces

[40] Beyond the above-mentioned notion of referent accessibility or salience (Ward et al., 1991; Gundel et al., 1993), other researchers have focused on related information-structural concepts like topichood (e.g., Kertz, 2013).

an IDENTITY relation between the elided material and its linguistic antecedent; and referential theories, which assume that ellipsis is governed by the same set of mechanisms that enable other forms of discourse reference. This distinction corresponds to H&S's (Hankamer & Sag, 1976; Sag & Hankamer, 1984) architectural distinction between "surface" anaphors, which depend directly on their linguistic antecedent, and "deep" anaphors, which refer to entities in interlocutors' shared model of the discourse and only indirectly depend on their antecedents. While H&S argued based on diagnostic properties of discourse reference that ellipsis is architecturally distinct (i.e., governed by a fundamentally distinct part of the language architecture), I extended their analysis in two ways, which led me to a different conclusion. First, I considered a broader range of diagnostic properties and found that both VP-ellipsis and sluicing behave exactly as one would expect under a referential theory of ellipsis. Those parallels are particularly striking given the fact that Gapping does not exhibit any of them, providing a baseline for what a purely syntactically governed form of ellipsis (or, in H&S's terms, a true "surface" anaphor) might behave like. Secondly, I reviewed patterns of inferential reference resolution with respect to non-elliptical forms of reference and found that they do exhibit the kind of sensitivity to morphosyntactic properties of their antecedent that H&S argued was unique to ellipsis.

In this section, I will compare referential theories and IDENTITY theories of ellipsis directly. First, I will consider cases involving argument-structure mismatches, which have historically played a central role in debates between IDENTITY theorists and advocates of referential theories (Ross, 1969; Chung et al., 1995; Kehler, 2000; Merchant, 2001; Arregui et al., 2006; Chung, 2006; Kim, Kobele, Runner, & Hale, 2011; Chung, 2013; Frazier, 2013; Kertz, 2013; Merchant, 2013b; Kim & Runner, 2018; Poppels & Kehler, 2019). As we will see, the empirical picture in this domain is complex and provides arguments for and against both approaches. I will then consider a phenomenon known as "connectivity effects." Whereas the gradient status of argument-structure mismatches has been interpreted by both camps as supporting their theory, connectivity effects are widely considered the strongest evidence in favor of IDENTITY theo-

ries (Chung et al., 2011; Lipták, 2015; Messick, Saab, & Vicente, 2016). I will consider both IDENTITY-based accounts of the facts as well as referential explanations and argue that they both capture the data equally well (albeit in fundamentally different ways), undermining the notion that connectivity effects favor IDENTITY theories. Finally, I will compare referential and IDENTITY theories with respect to theoretical parsimony considerations and outline the contributions of the remaining chapters against the theoretical background I am painting in this one.

2.3.1 Argument-structure mismatches

At first glance, VP-ellipsis appears to be infelicitous whenever the ellipsis clause and the antecedent clause involve different syntactic configurations, as shown in (88).[41]

(88) a. The problem was looked into by John, and Bob did #(look into the problem), too.
(Kehler, 2000, ex. 34)

b. Even if you want me to shut up, you can't #(shut me up).

This pattern receives a straightforward explanation from IDENTITY theories: assuming that differences in argument structure are rooted in distinct lexical items (Hale & Keyser, 1993), any theory that prohibits the ellipsis of lexical items not provided by the antecedent will correctly rule out those examples (e.g., Chung, 2006, 2013; Rudin, 2019). Furthermore, Rudin's (2019) structure-matching condition provides an additional constraint against argument-structure alternations by barring the word-order differences they incur.

While it is often assumed that the existence of such mismatch effects is problematic for referential theories (e.g., Arregui et al., 2006; Lipták, 2015), we have seen in Section 2.2 that

[41] Note that matched variants of these examples are acceptable:

(i) John looked into the problem, and Bob did (look into the problem), too.

(ii) Even if you want to shut me up, you can't (shut me up).

the existence of mismatch penalties is perfectly consistent with the mechanisms that support discourse reference, especially when they are found to exhibit gradience and vary across contexts. Indeed, that appears to be the case here as well: the following examples are appreciably more acceptable than the ones in (88), even though they involve the same kinds of mismatches.

(89) a. This problem was to have been looked into, but obviously nobody did (look into the problem). (Kehler, 2000, uttered by Vincent Della Pietra in conversation)

b. And I know that as much as some of you might want me to (shut up), it's 2018 and I'm a woman so you cannot shut me up.[42]

This state of affairs has motivated a number of experimental studies with the goal of explaining the gradience associated with constructional mismatches (Arregui et al., 2006; Kim et al., 2011; Kim & Runner, 2018; Poppels & Kehler, 2019). IDENTITY theorists typically aim to explain it as a by-product of processing mechanisms (e.g., following Arregui et al.'s Recycling Hypothesis) or other factors that operate above and beyond the grammatical constraints on ellipsis in order to maintain a binary grammatical classification despite the gradience in acceptability (see Section 2.1.2 and Chapter 3 for an in-depth discussion of this approach). Referential approaches, on the other hand, straightforwardly predict the possibility of gradience without the need for additional assumptions: whenever the meaning of the ellipsis site is not reducible to the antecedent-provided meaning, the intended referent must be inferred, which is often associated with reduced acceptability in utterances involving ellipsis as well as non-elliptical forms of discourse reference. Whether or not and to what extend inferential reference resolution incurs a penalty depends on a variety of factors, and two such factors have been found to play a key role in determining the gradient acceptability of voice-mismatched ellipsis (as well as other types of mismatches): coherence establishment (Kehler, 1993b, 2000), and information structure

[42]Michelle Wolf during the 2018 White House Correspondents Dinner, available at the time of writing at https://youtu.be/L8IYPnnsYJw?t=2m26s.

(Kertz, 2008, 2013; Miller & Pullum, 2013; Miller & Hemforth, 2014).

While the source of gradient acceptability patterns associated with cases of syntactic mismatch remains controversial, they provide *prima facia* support for referential theories because they predict their existence without additional assumptions. However, most experimental research in this area has focused on VP-ellipsis,[43] raising the question whether sluicing exhibits a similar kind of gradience. The significance of constructional mismatches to theories of sluicing has long been acknowledged (Levin, 1982; Merchant, 2008; Tanaka, 2011b; Merchant, 2013b; Lipták, 2015; Rudin, 2019), and many authors cite infelicitous examples of argument-structure mismatches in defense of IDENTITY theories of ellipsis (e.g., Merchant, 2005; Chung, 2006, 2013; Merchant, 2013b; Rudin, 2019). However, there has been little experimental work aimed at confirming the categorical nature of these effects. Chapter 5 addresses this issue by reporting two experiments investigating sluicing under mismatches due to "tough movement" and the active/passive voice alternation.

2.3.2 Connectivity effects

One key point of divergence between IDENTITY accounts and referential theories of ellipsis concerns the content of the ellipsis site. According to IDENTITY theories, it contains fully formed syntactic structure that simply remains unpronounced under ellipsis, whereas referential theories assume that it merely contains a phonologically null pro-form. As Merchant (2019) points out, questions about the nature of unpronounced linguistic material can only be addressed indirectly:

> Detecting and arguing for such 'missing' structures is analogous to searching for and determining the properties of a black hole: one can tell it's there only by its effects on surrounding material. The logic of the hunt for elided structure is similar.

[43] But see SanPietro, Merchant, and Xiang (2012) for an experimental study of sluicing, which will be discussed in more detail in Chapter 5.

> If one finds effects that seem to be due to missing material, there is an argument that such structure exists. (Merchant, 2019, p. 25)

It is therefore unsurprising that so-called "connectivity effects," exemplified in (90)-(92), have received ample attention in the literature, going back as far as Ross's (1969) famous observations regarding Case connectivity constraints on the distribution of sluicing in German.

(90) a. Sie werden jemanden.ACC entlassen, aber keiner weiß, {wen.ACC | #wem.DAT}.
They will someone.ACC fire, but nobody knows, {who.ACC | #who.DAT}.
'They will fire someone but nobody knows who.'

b. Sie werden jemandem.DAT kündigen, aber keiner weiß, {#wen.ACC | wem.DAT}.
They will someone.DAT fire, but nobody knows, {#who.ACC | who.DAT}.
'They will fire someone but nobody knows who.'

(91) a. Beth's wedding was in Bond Chapel, and Rachel's {was | *were} in Rockefeller Chapel. (adapted from Merchant, 2019, ex. 37-38)

b. Beth's nuptials were in Bond Chapel, and Rachel's {*was | were} in Rockefeller Chapel. (adapted from Merchant, 2019, ex. 37-38)

(92) a. Mike was supposed to give a eulogy at the funeral, but he {didn't | *wasn't}.

b. Mike was supposed to be at the funeral, but he {*didn't | wasn't}.

In each case, the ellipsis remnants appear to depend on (i.e., be "connected" to) the antecedent in some way: in (90) the remnant wh-phrase exhibits the same case marking as its correlate

jemanden/m 'someone' in the antecedent clause; in (91) the verb in the ellipsis clause agrees in number with the antecedent NP (*wedding* vs. *nuptials*); and in (92) the choice of the remnant auxiliary depends on the antecedent VP.

Connectivity effects follow straightforwardly from the core assumption behind IDENTITY theories of ellipsis: if the ellipsis site contains fully formed syntactic material that is identical to its antecedent, connectivity constraints on the ellipsis remnants can be analyzed as internal to the ellipsis clause, as illustrated in (93). For example, the elided verb *entlassen/kündigen* 'fire' assigns Case to the sluicing remnants *wen*.ACC/*wem*.DAT 'who' in (93a-b) ; the elided NP *wedding/nuptials* agrees in number with the verb *was/were* in (93c-d); and the elided phrase in (93e-f) constrains the distribution of the remnant auxiliary under VP-ellipsis.

(93) a. IDENTITY constraint / Case assignment

…jemanden.ACC entlassen, …{wen.ACC | # wem.DAT} ~~sie~~ ~~entlassen~~ ~~werden~~.

…someone.ACC fire, …{who.ACC | # who.DAT} ~~they~~ ~~fire~~ ~~will~~.

'They will fire someone but no one knows who.'

b. IDENTITY constraint / Case assignment

…jemandem.DAT kündigen, …{# wen.ACC | wem.DAT} ~~sie~~ ~~kündigen~~ ~~werden~~.

…someone.ACC fire, …{# who.ACC | who.DAT} ~~they~~ ~~fire~~ ~~will~~.

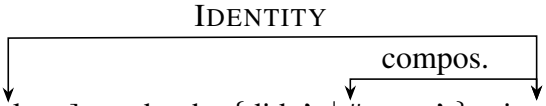
'They will fire someone but no one knows who.'

c. IDENTITY constraint / Agree

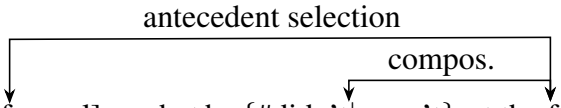
Beth's wedding was in Bond Chapel, and Rachel's ~~wedding~~ {was | # were} in...

d. IDENTITY constraint / Agree

Beth's nuptials were in Bond Chapel, and Rachel's ~~nuptials~~ {# was | were} in...



e. Mike was supposed to [give a eulogy].DO, but he {didn't | # wasn't} ~~give one~~.



f. Mike was supposed to [be at the funeral].BE, but he {#didn't|wasn't} ~~at the funeral~~.

The fact that the unelided counterparts of the elided utterances in (93) exhibit the same distribution, as shown in (94), makes the IDENTITY-based explanation particularly compelling (Van Craenenbroeck & Merchant, 2013; Lipták, 2015; Merchant, 2019): if IDENTITY theories are on the right track, elided and unelided utterances are syntactically indistinguishable and are therefore expected to pattern together with respect to Case assignment, number agreement, and so forth.

(94) a. Sie werden jemanden.ACC entlassen, aber keiner weiß, {wen.ACC | # wem.DAT} sie entlassen werden.

b. Sie werden jemandem.DAT kündigen, aber keiner weiß, {# wen.ACC | wem.DAT} sie kündigen werden.

c. Beth's wedding was in Bond Chapel, and Rachel's wedding {was | # were} in Rockefeller Chapel.

d. Beth's nuptials were in Bond Chapel, and Rachel's nuptials {# was | were} in Rockefeller Chapel.

e. Mike was supposed to give a eulogy at the funeral, but he {didn't | # wasn't} give a eulogy.

f. Mike was supposed to be at the funeral, but he {# didn't | wasn't} at the funeral.

As for the IDENTITY relation that is required for this explanation to go through, notice that any lexico-syntactic condition applied to the relevant lexical item (whichever syntactic node is responsible for Case assignment, number agreement, etc.) is sufficient (Chung, 2006; Merchant, 2013a, 2013b; Rudin, 2019). With respect to Case, this lexical identity requirement is perhaps most explicit in Chung's (2013) "Case condition":

> If the interrogative [sluicing remnant] is a DP, it must be Case-licensed in the ellipsis site by a head identical to the corresponding head in the antecedent clause.
>
> (Chung, 2013, p. 30)

It is worth noting, however, that purely semantic IDENTITY theories (e.g., Merchant, 2001), are not sufficient for capturing the pattern in (93): as Merchant (2019) points out, *nuptials* and *wedding* are synonymous and differ only with respect to the grammatical number feature that produces the connectivity effect. In order for the IDENTITY condition to prevent replacing one with the other under ellipsis, it must therefore be sensitive to this feature and cannot be defined in purely semantic terms. Likewise, *entlassen* and *kündigen* both mean 'fire' and differ only with respect to Case assignment.[44] If the two were allowed to vary under ellipsis, we wouldn't expect to see the connectivity effect we see in (93a-b).

Because connectivity effects "fall out" naturally from the assumption that the Case-assigning elements in the ellipsis and antecedent clauses are linked via the IDENTITY condition, they are often cited as evidence that favors IDENTITY theories over referential theories (Chung et al., 2011; Lipták, 2015; Messick et al., 2016), but that view is not universal among IDENTITY theorists. For example, Merchant (2019) argues that both IDENTITY theories and referential theories are, in principle, consistent with the existence of connectivity effects, since both assume *some* amount of silent linguistic structure at the ellipsis site. He classifies both types of theories as "structural approaches," and argues that connectivity effects merely serve as evidence against

[44] Indeed, these two verbs are arguably also equivalent in terms of register and lexical frequency, which addresses the potential concern that the *wedding*/*nuptials* pair is not fully matched pragmatically.

entirely *non*-structural approaches (e.g., Culicover & Jackendoff, 2005, 2012). Indeed, Ginzburg (1992), Jäger (2001, 2005), and Barker (2013) all argue that Case connectivity can be derived from referential analyses of sluicing, and, as we will see next, that explanation naturally extends to other connectivity effects and types of ellipsis as well.

Consider the following examples in which the morphosyntactic gender or number of the antecedent determines which pro-form can subsequently be used to refer to the entity it introduces into the common ground.

(95) a. Esta mesa.FEM me encanta. —No {la.FEM | # lo.MASC} veo.
This table.FEM me delights. —Not {her.FEM | # him.MASC} see.
'I like this table. —I don't see it.'

b. Este partido.FEM me encontó. —No {# la.FEM | lo.MASC} ví.
This match.FEM me delighted. —Not {# her.FEM | him.MASC} saw.
'I liked this match. —I didn't see it.'

(96) a. I haven't worn these pants in years, I even forgot {they | # it} existed. ≈ (46c)
b. I haven't worn this shirt in years, I even forgot {# they | it} existed.

We can represent these antecedent-selection constraints as shown in (97), where subscripts on referring expressions indicate what type of antecedent they require.[45]

(97) a. Esta mesa.FEM me encanta. —No {la$_{fem}$ | # lo$_{masc}$} veo.
antecedent selection

b. Este partido.MASC me encontó. —No {# la$_{fem}$ | lo$_{masc}$} ví.
antecedent selection

[45] While we focus notationally on a subset of properties that are relevant in any particular case, note that other properties may matter as well. For example, in (96b) *she* would be infelicitous even though it matches the antecedent in number, presumably because of a mismatch in gender or animacy.

antecedent selection

c. I haven't worn these pants.PL in years, I even forgot {they$_{pl}$ | # it$_{sg}$} existed.

antecedent selection

d. I haven't worn this shirt.SG in years, I even forgot {# they$_{pl}$ | it$_{sg}$} existed.

It is important to emphasize at this point that none of the examples in (97) involve ellipsis and thus demonstrate that the underlying mechanism that produces antecedent-selection effects (whatever it may be) is needed independently of ellipsis. Nonetheless, the same machinery can be leveraged to explain connectivity constraints under ellipsis, which is an idea that goes back at least as far as Ginzburg (1992) and has subsequently been formalized by Jäger (2001, 2005) and Barker (2013). According to that analysis, connectivity effects arise as the result of two constraints, as illustrated in (98): a local constraint that ensures that the pro-form at the ellipsis site can compose with the ellipsis remnants, and an antecedent-selection constraint that "connects" the ellipsis site to its antecedent in terms of the relevant properties.

(98) a. Sie werden jemanden.ACC entlassen, AKW[46] {wen.ACC | # wem.DAT} σ_{acc} .[47]

antecedent selection composability

b. Sie werden jemandem.DAT kündigen, AKW {# wen.ACC | wem.DAT} σ_{dat} .

antecedent selection composability

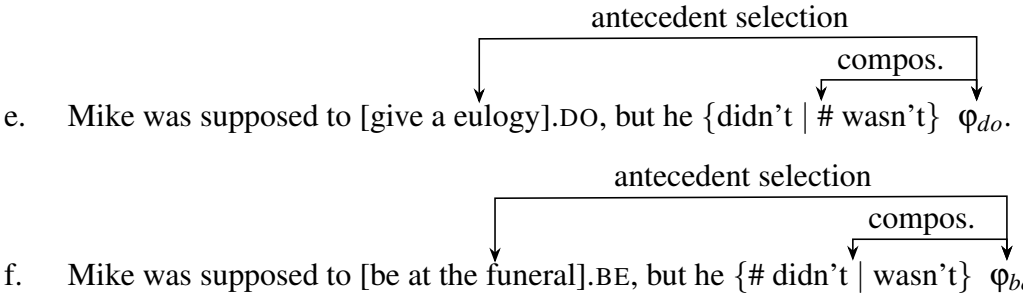
c. Beth's wedding was …, and Rachel's η_{sg} {# was | were} ….

antecedent selection composability

d. Beth's nuptials were …, and Rachel's η_{pl} {# was | were} ….

[46]I use AKW as a short hand for *aber keiner weiß* 'but nobody knows.'

[47]For readability, I omit the gloss, which is identical to the one in (93a-b).



The local "composability" constraint that ensures that the pro-form can compose with the ellipsis remnants applies equally to unelided variants,[48] and it is, in fact, entirely analogous to the local constraint IDENTITY-based explanations invoke as described above.

The explanations differ with respect to the second constraint, however, and this is where their respective architectural assumptions come into play: whereas the IDENTITY condition is a made-for-purpose constraint that only applies to ellipsis, the antecedent-selection constraint is independently motivated by the behavior of non-elliptical forms of discourse reference. In that sense, the referential explanation is more parsimonious because it avoids stipulating ellipsis-specific machinery, whereas the IDENTITY-based explanation misses the generalization that connectivity effects are not only associated with ellipsis, but also with non-elliptical referential expressions, as illustrated in (97). On the other hand, the referential explanation requires the proliferation of ellipsis pro-forms in the lexicon: we must stipulate a separate pro-form for each relevant set of antecedent properties (e.g., σ_{dat}, σ_{acc}, σ_{nom}, etc.).[49] While introducing multiple pro-forms that perform the same general function and differ only with respect to certain antecedent-selection criteria may not be problematic in principle (after all, the assumption that *he* and *she* are separate lexical items is commonplace), doing so for phonologically null elements may appear theoretically "expensive" in its own way. Nonetheless, the common assumption that

[48] See Jäger (2001, 2005) and Barker (2013) for a formalization of this constraint within the framework of Type Logical Grammar.

[49] I am intentionally glossing over the fact that Jäger (2001, 2005) analyzes the remnants themselves as anaphoric. While this approach avoids stipulating pro-forms at the ellipsis site, it nonetheless requires the introduction of an anaphoric variant of each lexical item that can serve as an ellipsis remnant.

connectivity effects overwhelmingly favor IDENTITY theories over referential theories of ellipsis is incorrect: both types of accounts provide empirically adequate explanations that differ only in the core architectural assumptions they are built upon, and raise certain concerns with respect to theoretical parsimony. In the following, final section of Chapter 2, I will expand on issues of theoretical parsimony beyond connectivity.

2.3.3 Overcoming the "IDENTITY crisis"

One of the most fundamental differences between IDENTITY theories and referential theories of ellipsis is that the former propose an ellipsis-specific constraint whereas the latter aim to explain ellipsis in terms of the independently motivated mechanisms behind discourse reference. Consider, as an example, Merchant's (2001) e-GIVENness, which remains one of the most influential proposals to date. While it is based on the notion of GIVENness that Schwarzschild (1999) used to explain patterns of focus marking and pitch accent, it crucially goes beyond this independently motivated one-way entailment requirement by adding an ellipsis-specific "reverse entailment" condition (hence the "e" in e-GIVENness). Not only does this condition by definition only apply to the use of ellipsis, there is to my knowledge no other linguistic phenomenon that requires a "downstream" element to entail its antecedent. The same is true of virtually all other IDENTITY-based proposals (e.g., Chung, 2006; Elbourne, 2008; Chung, 2013; Rudin, 2019): they all introduce conditions and mechanisms that are specifically designed to handle ellipsis, instead of recruiting mechanisms that are independently needed to explain other phenomena. As I argued throughout this chapter, the assumption that ellipsis is governed by *sui generis* mechanisms follows from H&S's (Hankamer & Sag, 1976; Sag & Hankamer, 1984) conclusion that elliptical phenomena are architecturally distinct from other forms of context dependency. By contrast, referential theories of ellipsis are based on the fundamental architectural assumption that ellipsis is governed by independently motivated machinery, namely the system of discourse reference, which makes referential theories inherently more parsimonious: since the theoretical

constructs leveraged by referential theories of ellipsis—such as referent accessibility, knowledge-driven inference mechanisms, etc.—are needed to explain discourse reference phenomena beyond ellipsis, pivoting away from the view that elliptical utterances are referential in nature does not actually simplify the overall picture of how (context-dependent) language works.

Beyond those considerations of theoretical parsimony, referential theories of ellipsis also have an empirical advantage over IDENTITY theories: since referent accessibility and interlocutors' beliefs about the extent of their mutual knowledge are inherently gradient in nature, the use of ellipsis is expected to be gradiently felicitous, especially when the intended referent has to be inferred because it is not reducible to the denotation of the antecedent. IDENTITY theories, on the other hand, predict a categorical distribution: either ellipsis is grammatical because its use satisfies the IDENTITY condition, or it is ungrammatical because it does not. In order to derive gradience in acceptability, IDENTITY theories typically appeal to theory-external mechanisms, such as the Recycling Hypothesis. While invoking this type of competence/performance distinction to explain gradient acceptability patterns is a legitimate theoretical stance in principle, the biggest empirical challenge for IDENTITY theories lies in the fact that there is a plethora of acceptable mismatches that elude even the most recent definitions of IDENTITY that have emerged after decades of fine-tuning (see Section 2.1), and that are likewise outside the reach of Recycling-type approaches (Poppels & Kehler, 2018; Frazier & Duff, 2019; Poppels & Kehler, 2019, see also Chapters 4 and 6).

This "IDENTITY crisis" is particularly severe because the fine-tuning of theory-internal and -external parameters in order to "capture" more observations undermines the explanatory value of IDENTITY theories in a way that is analogous to Ptolemaic epicycles. The Ptolemaic world view was based on the axiomatic assumption that the Earth is at the center of the solar system and that all other planetary objects known at the time revolve around it. To square this theory with apparently inconsistent planetary movement as observed from Earth (now known as "apparent retrograde motion"), Ptolemy stipulated that planets follow the paths of epicycles, i.e., circles on circles

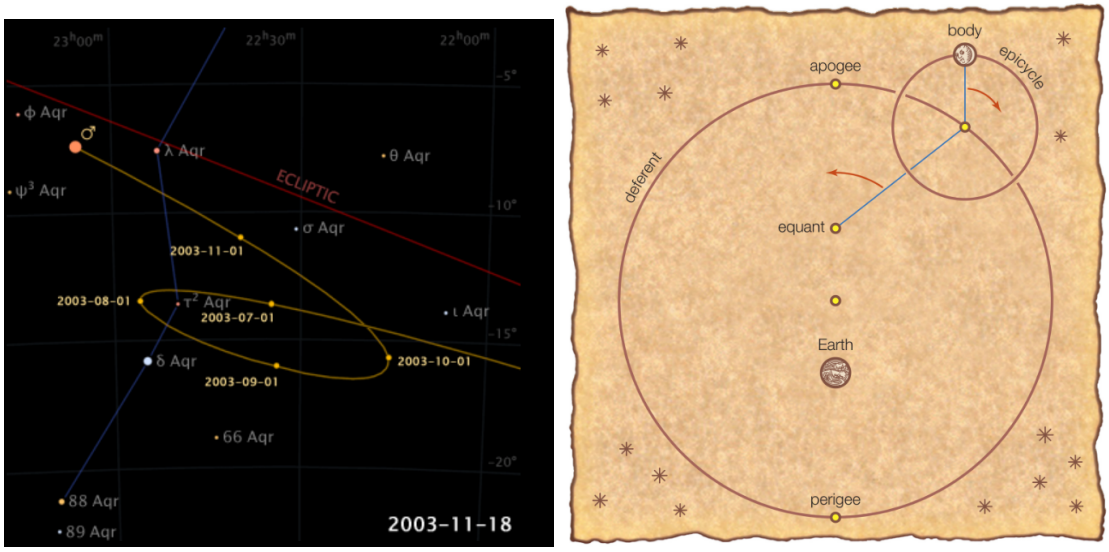


Figure 2.1: Left: Apparent retrograde motion of Mars (yellow) as observed from Earth in 2003 (created by Eugene Alvin Villar as part of the Philip Greenspun illustration project and released under the CC BY-SA 4.0 license). Right: Schematic representation of Ptolemy's geocentric model of the universe.[51]

(see Figure 2.1). With enough epicycles in the right places, Ptolemy's model of planetary motion achieved impressive accuracy with respect to the observations made from Earth,[50] despite the fact that it was based on fundamentally misguided architectural assumptions about the universe. In other words, Ptolemy *captured* observations by adding parameters, but his model fundamentally failed to *explain* the observations. My concern about IDENTITY theories of ellipsis is that incrementally fine-tuning the definition of IDENTITY and exploiting theory-external degrees of freedom (e.g., about the nature of mismatching elements or the competence/performance distinction; see Sections 2.1.2 and 2.1.3)—in the service of maintaining the assumption that ellipsis is governed by IDENTITY—will similarly fail to *explain* ellipsis, even if it succeeds in *approximating* its distribution increasingly well.

To overcome this "IDENTITY crisis," it is imperative that we reconsider the fundamental

[50] With the discoveries of Joseph Fourier, it was proven centuries later that the epicycle system used by Ptolemy and many other ancient astronomers tracks a generalizable approximation system that can describe any arbitrary curve given enough parameters.

[51] Taken from the Encyclopedia Britannica at https://www.britannica.com/science/Ptolemaic-system#/media/1/482079/60364.

assumption at its core by asking whether ellipsis really is governed by some yet-to-be-defined IDENTITY condition. Instead of asking how to capture the distribution of ellipsis, I followed H&S's lead in asking whether ellipsis behaves like other context-sensitive linguistic expressions with respect to a range of diagnostic properties. Motivated by the parallels that emerge from this high-level comparison, analyzing VP-ellipsis and sluicing as referential expressions provides a way of avoiding the epicycle dynamic that IDENTITY theories are vulnerable to: since the explanatory constructs available to referential theorists equally apply to non-elliptical discourse reference, we cannot fine-tune theoretical parameters based on ellipsis data alone. That being said, referential theories of ellipsis do face various theoretical and empirical challenges that must be addressed in future research.

Two central questions for theories of discourse reference, which are the focus of the remaining chapters of this thesis, are about the notion of referent accessibility and, relatedly, why inferential reference resolution succeeds in some cases but not others. Why, for example, is the pronoun in the Partee's marbles example discussed above and repeated in (99) infelicitous despite the fact that the intended referent is highly salient and easily recognizable as mutually known by the comprehender? Similarly, why is the Hartman example in (100a) impossible while the example in (100b) is perfectly felicitous despite involving a similar mismatch between relational opposites?

(99) I dropped ten marbles and only found 9 of them. # It's under the sofa.

(100) a. Billy won against someone at chess, and Susan did #(lose to someone), too. (Hartman, 2009)

b. Mary's boyfriend gave her his school picture, just like all schoolboys do (give their girlfriends their school picture). (Kehler, 2002a)

These questions are not trivial. In fact, some of them may well be *AI complete* (i.e., completely

answerable only in the context of a complete theory of human intelligence) because the mechanisms behind discourse reference interface with other AI complete cognitive mechanisms, such as interlocutors' conceptual knowledge and the theory-of-mind mechanisms involved in establishing and coordinating mutual knowledge. While the difficulty of answering these questions is sometimes raised as an argument against referential theories of ellipsis, it is important to emphasize that these questions exist in the context of non-elliptical reference regardless of which theory of ellipsis we adopt.

This thesis thus has two goals: to continue to make the case for referential theories of ellipsis; and to push for explanatory theories of how elliptical reference works, what factors govern inferential ellipsis resolution and why. Both of these goals are served by focusing on cases involving various types of mismatch, since they are both inherently challenging for IDENTITY theories while at the same time exposing the inferential mechanisms that enable them according to referential theories of ellipsis. Chapter 3 begins by considering sources of gradience associated with voice-mismatched VP-ellipsis and argues based on a series of experiments that the data are beyond the explanatory reach of both "pure" IDENTITY theories as well as processing theories like the Recycling Hypothesis. Chapter 4 then explores cases of inferential VP-ellipsis like (101) in which the ellipsis clause acquires a meaning that is related, but not identical, to the event denoted by the antecedent VP. I will argue that these examples, too, elude IDENTITY-based explanations while patterning closely with non-elliptical reference to inferred entities.

(101) a. A: Can I borrow your textbook over the weekend?
B: I can't (lend it to you), I'll need it myself.

b. Spectator: Can I see that card trick one more time?
Magician: I can't (show it to you again), sorry.

Chapter 5 then turns to sluicing and presents experimental evidence for acceptable argument-structure mismatches based on "tough movement" and the passive/active voice alternation, as

shown in (102). The results from these experiments are inconsistent with several formulations of IDENTITY (e.g., Merchant, 2013b; Rudin, 2019) and raise novel questions for referential theories of sluicing.

(102) a. Break fluid is easy to replace if you know how (to replace break fluid).

b. The problem hasn't been solved because no one knows how (to solve it).

Finally, Chapter 6 considers even more disruptive examples of sluicing that involve various types of mismatch, exemplified in (103) and (104).

(103) Fan: Can I get a few autographs?

Manager: Sure, how many (do you want/need)?

(104) Regarding Trump's impeachment, the only question is...

a. ...when (he will be impeached).

b. ...why #(he will be impeached).

c. ...who ##(he will be impeached by).

Some of these examples are highly acceptable, while others exhibit a tremendous amount of gradience, raising questions for IDENTITY theories and referential theories alike.

Chapter 3

Evaluating the Recycling Hypothesis

Recall from Chapter 2 that ellipsis theories can be divided into two camps based on their core assumptions about which part of the language architecture governs ellipsis: according to IDENTITY theories, ellipsis is grammatical only if the linguistic context provides an antecedent that is identical to the elided material; whereas referential theories posit that ellipsis is a form of discourse reference and requires that the intended referent be in common ground and sufficiently salient or else inferrable from the context of utterance. Given this theoretical picture, it is unsurprising that cases involving some degree of mismatch between the elided material and its antecedent have received particular attention in the literature, since such cases speak to the core assumptions behind both types of theory. For IDENTITY theories, mismatch cases represent potential counterexamples that inform the definition of IDENTITY that is required to approximate the distribution of ellipsis. According to referential theories, on the other hand, mismatch cases require some degree of inference since in such cases the intended referent cannot be reduced to the meaning provided by the antecedent-provided meaning; they are therefore potentially informative about the factors that enable and constrain inferential ellipsis resolution.

This chapter, which contains material published in Poppels and Kehler (2019), focuses on the use of VP-ellipsis when the ellipsis clause and its antecedent differ in Voice, as illustrated in

(105), repeated from Chapter 2.

(105) a. This problem was looked into by John, and Bob did #(look into the problem) too.

b. This problem was to have been looked into, but obviously nobody did (look into the problem). (Vincent Della Pietra, in conversation, cited in Kehler, 1993b)

Examples like (105a) suggest that voice mismatches are associated with reduced acceptability, in line with the predictions of (syntactic) IDENTITY theories (Sag, 1976; Chung, 2006, 2013; Rudin, 2019, and many others).[1] The challenge for referential analyses is to explain this unacceptability given that a meaning corresponding to the elided VP may seem to have been made available by the meaning of the antecedent clause. Example (105b), on the other hand, is widely judged to be acceptable, despite it having the same passive-active mismatch that characterizes (105a). Such examples hence challenge syntactic analyses, since the required VP is not available in this case either.[2]

The issues surrounding the status of examples that involve mismatch have inspired a considerable amount of experimental work that has sought to obtain more fine-grained measurements of acceptability, generally by utilizing acceptability rating tasks (Arregui et al., 2006; Kim et al., 2011; SanPietro et al., 2012; Kertz, 2013; Kim & Runner, 2018, among others). This research has uncovered two significant patterns. The first is that cases involving mismatch such as (106b) are reliably judged to be less acceptable than paired variants in which the voice is matched as in (106a).

(106) a. The judge read the report first, and then the lawyer did (read the report) too.

[1] See Merchant (2013b) for a syntactic IDENTITY analysis that selectively allows voice mismatches for VP-ellipsis while prohibiting them for types of ellipsis that target larger constituents, such as sluicing or fragment answers. I discuss this approach in more detail in Chapter 5, which presents experimental evidence that challenges Merchant's analysis.

[2] There are a variety of analyses on offer for reconciling these data (Kehler, 2002b; Kim et al., 2011; Grant et al., 2012; Kertz, 2013), the details of which will not concern us.

[match]

b. The report was first read by the judge, and then the lawyer did #(read the report) too. [mismatch]

The second pattern, which is in fact the primary concern of the work presented here, is the existence of differences in acceptability between different *types* of voice mismatches. Specifically, passive-voice VP-ellipsis with active antecedents, as in (107b), tends to be less acceptable than active voice VP-ellipsis with passive antecedents, as in (107a).

(107) a. The report was first read by the judge, and then the lawyer did #(read the report) too. = (106b) `[P -> A]`

b. The judge read the report first, and then the confession was #(read by the judge). `[A -> P]`

This finding, which was first reported by Arregui et al. (2006, Experiment 5), has been replicated in several subsequent studies (Kim et al., 2011; Parker, 2017; Xiang & Klafka, 2018).[3] We henceforth refer to this pattern as the MISMATCH ASYMMETRY. This finding gives rise to an immediate question: Does the Mismatch Asymmetry result from the linguistic properties of VP-ellipsis, and hence require a linguistic explanation, or does it reflect a fact about processing, external to the theory of VP-ellipsis itself? Pinning down the locus of the phenomenon is important for understanding both the linguistic constraints that govern the use of ellipsis as well as the mechanics of the interpretation process that comprehenders utilize to recover its meaning.

Several recent works have argued for a processing-based explanation, specifically based on the behavior of memory (Arregui et al., 2006; Parker, 2017; Xiang & Klafka, 2018). Here we focus specifically on the analysis of Arregui et al. (2006) (see also Frazier 2013), who explain the mismatch asymmetry by appeal to a processing theory known as the Recycling Hypothesis

[3]Note however that Kim et al. (2011) failed to find a reliable effect; see their Experiment 1.

(RH; see Section 2.1.3). The RH has two components: (i) a grammatical constraint on the use of VP-ellipsis, which requires syntactic IDENTITY between the elided material and its antecedent; and (ii) a processing theory to explain any residual variation in acceptability when the grammar predicts ungrammaticality. Given their notion of syntactic IDENTITY, voice-mismatched VP-ellipsis is categorically ruled out as ungrammatical, and as a result `[P -> A]` and `[A -> P]` mismatches are predicted to be equally unacceptable as far as the grammar is concerned. However, whenever the sentence processor is faced with a grammatical violation, it attempts to reanalyze past syntactic material and "recycle" it in a way that renders the input grammatical. In the case of ellipsis with non-identical antecedents, this Recycler is taken to reanalyze the existing antecedent and fashion an alternative antecedent that satisfies the IDENTITY constraint. The amount of work that the Recycler needs to carry out in order to repair an ellipsis is hypothesized to determine the relative level of acceptability of the passage in question, such that ungrammatical cases of ellipsis with non-identical antecedents may be perceived as relatively acceptable as long as an identical antecedent can be "recycled" from the existing one without a lot of effort.[4]

The asymmetry between `[P -> A]` and `[A -> P]` mismatches is explained as a by-product of the Recycling process with the help of an independently motivated auxiliary assumption, based on syntactic misremembering on the part of both the speaker and hearer. Specifically, the idea is that speakers, having selected a syntactic form among several to choose from in expressing a proposition, may not attend to the actual utterance they produced when planning the structure of a follow-on clause. In cases in which production involves a choice between systematic paraphrases such as active and passive variants of a clause, speakers may therefore inadvertently produce an ellipsis clause that doesn't match the voice of the antecedent clause, despite the fact that the result, according to the RH, is nonetheless ungrammatical. Furthermore, the RH also posits that a speaker's tendency to remember should be dependent on syntactic complexity: a

[4] As detailed in Section 2.1.3, this approach faces both theoretical concerns as well as empirical issues (e.g., Arregui et al., 2006). Here, we seek to expand this picture by evaluating the proposal against new experimental data, focusing specifically on the explanation it provides for the asymmetry between `[P -> A]` and `[A -> P]` mismatches.

more complex antecedent (e.g., a passive) should be more easily misremembered as a simpler one (e.g., an active) than the other way around. As a result, one expects to witness `[P -> A]` mismatches being produced more often than `[A -> P]` ones.

The same logic is then taken to apply to the hearer as well: previously heard passive clauses are more likely to be misremembered as having been in the active voice than previously heard active clauses are to be misremembered as having been in the passive (Mehler, 1963). Since the grammar licenses ellipsis only when the elided material is syntactically identical to the antecedent, the processing of the second clause requires the retrieval of the first clause from memory in order to evaluate whether the two are identical. When the to-be-retrieved clause is passive, there is some chance that it will be misremembered as active, resulting in an "illusory" `[A -> A]` match, which the authors refer to as an "illusion of grammaticality." This idea is illustrated in (108a), where *italic font* indicates a relatively noisy memory trace of the passive antecedent clause.

(108) a. *The report was first read by the judge* before the lawyer did too.
illusory `[A -> A]`

b. The judge read the report first before the confession was too. `[A -> P]`

Since active clauses are less prone to being misremembered as passive, mismatches as in (108b) are less likely to elicit such an illusion of grammaticality and are therefore, under the RH, predicted to receive a lower average acceptability rating. On this story, therefore, the effect is explained as a processing phenomenon, and hence requires no special accommodation within the theory of ellipsis itself.

We have now seen two sets of experimental data that potentially speak to the architectural differences between IDENTITY theories and referential theories of ellipsis, and a representative answer from advocates of syntactic analyses—the RH—that seeks to explain the effects by way of a processing model that lies external to the grammar of VP-ellipsis. The question now is what

sort of explanation could be offered under the posits of a referential theory. As we have seen in Chapter 2 (see also Kehler, 2017, 2019), the existence of mismatch effects is not necessarily surprising on referential theories, since it is well-known from research on entity-level reference that the linguistic form of an antecedent expression can affect the relative level of accessibility of the entity it denotes with respect to the hearer's mental model of the discourse. Applying this idea to the case of reference to eventualities, we observe that when a syntactic match exists as in (109a), a representation of the meaning of the referent has already been computed by way of the compositional semantic analysis of the VP, per (109b).

(109) a. John looked into the problem, and Bill did too.

b. [[VP]]: $\lambda x.look\_into(x, problem)$

c. The problem was looked into by John, and Bill did too.

d. [[VP]]: $\lambda x.look\_into(John, x)$

e. [[S]]: $look\_into(John, problem)$

That is to say, at the time that the ellipsis site is encountered, a representation of the referent will already be in the hearer's mental model of the discourse. This is not the case, however, when there is a syntactic mismatch as in (109c); here the compositionally-determined meaning of the VP, shown in (109d), is not the required one. Obtaining the necessary meaning (109b) will require an additional computation, e.g., the recovery of a lambda abstract from a representation of the meaning of the entire clause (109e).[5] So the idea that a modicum of additional discourse-level processing might be required to fashion a representation of the referent in the case of syntactic mismatches, under the presumption that VP-ellipsis presupposes that the representation is already available, could potentially explain their reduced acceptability.

Accounting for the Mismatch Asymmetry on a referential theory, however, appears to be more problematic. Here, the logic offered above to explain mismatch effects is of no

[5]For a procedure that resolves all VP-ellipses by way of such a calculation, see Dalrymple et al. (1991).

help: it should require no more work to fashion a representation for a passive VP from the meaning of a clause in the active voice than is required to fashion one for an active VP from the meaning of a clause in the passive voice. As such, the RH's memory-based explanation of the asymmetry is a potentially important argument for a syntactic analysis of VP-ellipsis. Because the misremembering phenomenon upon which the analysis is based is specific to syntactic (and not semantic) representations, proponents of the referential analysis have no similar story to tell.

This reasoning only goes through, of course, if the RH's analysis of the Mismatch Asymmetry is the correct one. The goal of this chapter is therefore to explore the source of the asymmetry, with particular attention to the predictions of the RH account on previously unexamined cases. Experiment 1 aims to replicate the Mismatch Asymmetry using stimuli adapted from the original Arregui et al. (2006) study, but to do so using a fuller paradigm that contains voice-matched control items that were not included in the original experiment. Whereas the results succeed in replicating the key finding of Arregui et al.'s study, they also reveal a penalty for passive ellipsis clauses even when the ellipsis clause is syntactically matched with the antecedent clause, suggesting the existence of a more general passive penalty for ellipsis clauses. Experiment 2 considers the question of whether the passive penalty might be independent of ellipsis *per se*, in part by examining the acceptability of unelided variants of the stimuli used in Experiment 1. Whereas the effect found in Experiment 1 was replicated for the elided versions, no such effect was found for the unelided variants, indicating that the passive penalty is specific to ellipsis clauses. Experiment 3 then provides the critical test of the RH analysis by examining cases that feature cataphoric VP-ellipsis. In such cases, the RH and the passive penalty hypothesis make opposing predictions: the passive penalty hypothesis predicts that mismatches with passive ellipsis clauses should remain worse than those with active ellipsis clauses, whereas the RH predicts that the judgments should reverse, since it is now the structure of the ellipsis clause, by virtue of occurring first, that is subject to misremembering. The results support the existence of a more general passive penalty for ellipsis clauses as opposed to a memory-based explanation. As

an ensemble, the results therefore suggest that neither memory-based explanations such as the RH nor any ellipsis-independent explanation is capable of accounting for the Mismatch Asymmetry. We conclude by discussing some ramifications of our results for the debate between syntactic and referential theories of VP-ellipsis, as well as a possible source of the effect.

3.1 Experiment 1

As described above, Arregui et al. (2006) found a Mismatch Asymmetry whereby `[A -> P]` mismatches in VP-ellipsis were subject to a greater acceptability penalty than were `[P -> A]` mismatches. However, no matched controls were included, which are necessary to confirm that the effects are specific to mismatched cases.[6] Because the RH explains the mismatch asymmetry as being a by-product of the Recycling process, and it explicitly bars the Recycler from being recruited unless there is a grammatical violation, one expects that the difference between active and passive clauses in terms of memory retrieval will have no effect on voice-*matched* VP-ellipsis. The purpose of Experiment 1 is thus twofold: to replicate the Mismatch Asymmetry that Arregui et al. found, and to include voice-matched controls to further examine the predictions of the analysis.

3.1.1 Methods

Stimuli

Twenty-four experimental items followed a 2x2 design, crossing two independent factors: whether the ellipsis clause and the antecedent MATCHED or MISMATCHED in voice, and whether

[6] Indeed, other recent work suggests that the effects might not be, although the evidence isn't unequivocal. For instance, Kim et al. (2011) found that cases of `[P -> P]` matched VP-ellipsis were rated as less acceptable that `[A -> A]` matched cases, but do not discuss the effect further. Parker (2018) and an unpublished study by Xiang and Klafka (2018) report similar effects for their stimuli. On the other hand, Kim and Runner (2018) report a significant interaction between mismatch and antecedent voice, suggesting that voice-matched `[A -> A]` and `[P -> P]` stimuli did not differ in acceptability, although the data analyzed included non-elliptical variants as well. Hence we seek to investigate the question ourselves, using variants of Arregui et al.'s own stimuli.

the ellipsis clause was ACTIVE or PASSIVE, as illustrated in (110)–(111).

(110) a. The judge read the report first, and then the lawyer did too. [A -> A]

b. The report was first read by the judge, and then the confession was too. [P -> P]

c. The report was first read by the judge, and then the lawyer did too. [P -> A]

d. The judge read the report first, and then the confession was too. [A -> P]

(111) a. The customer praised the dessert after the critic did already. [A -> A]

b. The dessert was praised by the customer after the appetizer was already. [P -> P]

c. The dessert was praised by the customer after the critic did already. [P -> A]

d. The customer praised the dessert after the appetizer was already. [A -> P]

The mismatch variants were identical to the stimuli used in Arregui et al. (2006, Experiment 5), setting aside the correction of a small number of typos and a few changes so as to ensure that all clauses were plausible and identical across item variants. The voice-matched variants were constructed by holding the ellipsis clauses of the mismatched variants constant and exchanging the antecedent clauses, leaving everything else unchanged. As with Arregui et al.'s stimuli, in half of the stimuli the antecedent and ellipsis clauses were conjoined with *and (then)...too* as in (110), and in half they were in a subordinating configuration using the connective *after* (111). The items were supplemented with 48 filler items exemplified in (112): 24 acceptable fillers and 24 unacceptable ones, half of each involving ellipsis and half not.

(112) a. The thief was arrested and his brother was as well. Acceptable, elliptical filler

b. A proof that God exists doesn't.[7] Unacceptable, elliptical filler

c. I can't hear the announcement but I don't care. Acceptable, non-elliptical filler

d. What did you meet a janitor that hates? Unacceptable, non-elliptical filler

[7]This particular item is due to Sag (1976), who points out that it appears to be ungrammatical.

These filler items were designed to establish clear upper and lower bounds in terms of acceptability. The non-elliptical fillers were also intended to distract participants from the purpose of the experiment.

Procedure

We recruited 30 participants via Amazon.com's Mechanical Turk, one of whom reported being a non-native English speaker and was excluded from all analyses. In a within-item and within-participant design, each participant was presented with exactly one variant of each of the 24 experimental items, which were presented in a random order and interspersed with the 48 filler items exemplified in (112). The materials were presented using the Ibex software for conducting psycholinguistic experiments online[8] and participants were instructed to rate each item in terms of its acceptability on a scale from 1-5, with a 5 rating meaning that "the sentence is perfectly acceptable in English and that you can imagine yourself or other native speakers saying it."

3.1.2 Predictions

We examine three predictions that are derived from the RH. First, we should find items in the MISMATCH condition to be degraded compared to their MATCHED counterparts, replicating the effect found in previous experiments. This prediction follows from the grammatical constraint of syntactic IDENTITY enforced by the RH and other syntactic analyses. Second, [A -> P] mismatches should be less acceptable than [P -> A] mismatches, per the Mismatch Asymmetry. Third, if the Mismatch Asymmetry is a by-product of the Recycling process, we expect to find no such difference between the two sets of voice-MATCHED items, since the syntactic IDENTITY condition is satisfied in those cases and hence the Recycler is not recruited.

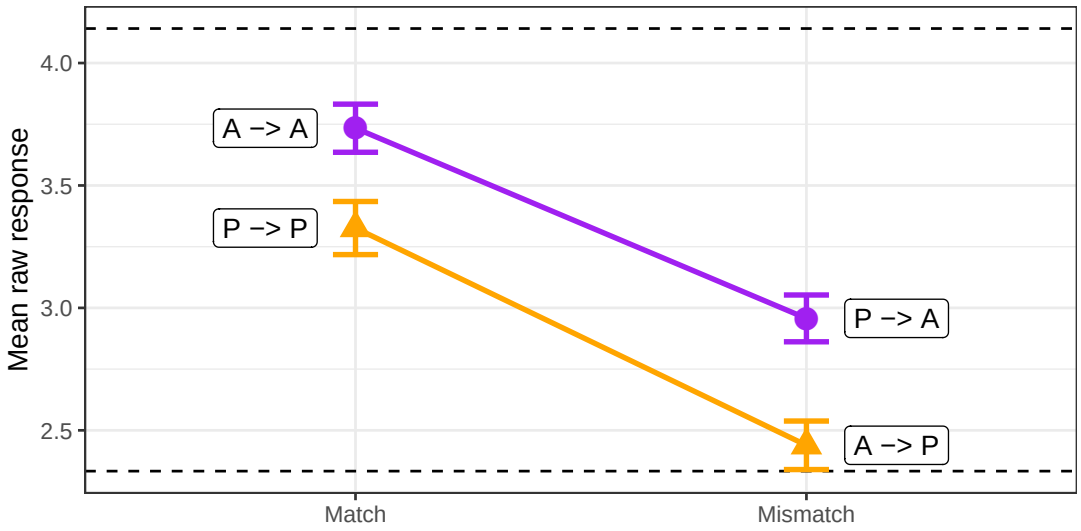


Figure 3.1: Results from Experiment 1. Dashed lines indicate mean acceptability of (un)acceptable elliptical fillers. Error bars show Standard Errors

3.1.3 Results

The results are summarized in Figure 3.1. While we reported a Gaussian mixed effects regression analysis in Poppels and Kehler (2019), we also fit a cumulative probit ("ordered category") model as reported below. To account for the discrete and bounded nature of the experimental rating scale, this type of statistical model treats raw ratings as categorizations based on a latent acceptability metric and jointly infers threshold values and the effects of predictors relative to those thresholds. In addition to sum-coded population-level ("fixed") effects of MATCH/MISMATCH, VOICE and their interaction, we specified the maximal group-level ("random") effect structure justified by the design (Barr, Levy, Scheepers, & Tily, 2013), including all item- and participant-specific intercepts and slopes.[9] We fit the model with weakly informative priors using the `brms` R package (R Development Core Team, 2009; Bürkner, 2017), and report the results in terms of the model's posterior distribution over population-level parameters: the effect size (Δ), i.e. the (probit-scaled) difference between condition means; the 95% Credible

[8] https://github.com/addrummond/ibex

[9] The complete formula was: response ~ mismatch + voice.ellipsis + mismatch:voice.ellipsis + (1 + mismatch*voice.ellipsis | subject) + (1 + mismatch*voice.ellipsis | item).

Interval (CI) around those estimates; and the probability given the model and data that the effect is below zero (or above, depending on the hypothesis), $P(\Delta < 0)$.

The data reveal a significant mismatch penalty whereby examples with syntactically mismatched antecedent and ellipsis clauses were judged to be worse than examples in which the clauses were syntactically matched ($\Delta = -0.49$, $CI(\Delta) = [-0.71, -0.28]$, $P(\Delta < 0) = 1$). Further, `[A -> P]` mismatches were less acceptable than `[P -> A]` mismatches, which is reflected in a main effect whereby items with passive ellipsis clauses were significantly degraded compared to those with active ellipsis clauses ($\Delta = -0.29$, $CI(\Delta) = [-0.47, -0.12]$, $P(\Delta < 0) = 1$). These two main effects were independent of each other: we found no evidence for an interaction between the two ($\Delta = -0.03$, $CI(\Delta) = [-0.17, 0.1]$, $P(\Delta < 0) = 0.7$).

3.1.4 Discussion

The goal of Experiment 1 was to evaluate the three predictions outlined in Section 3.1.2. Consistent with previous findings and the predictions of the RH, the results confirmed the first prediction, whereby syntactically mismatched cases of VP-ellipsis were reliably judged to be less acceptable than syntactically matched cases. The second prediction was also borne out: as expected, `[A -> P]` mismatches were less acceptable than `[P -> A]` mismatches, replicating the Mismatch Asymmetry effect identified by Arregui et al. The results failed to confirm the third prediction of the RH, however, according to which there should be no analogous effect in syntactically matched cases. Instead, a parallel effect was in fact found, whereby `[P -> P]` matches were rated as less acceptable than `[A -> A]` matches, with no interaction.

The results therefore cast doubt on the RH's memory-based explanation of the Mismatch Asymmetry. Instead, they suggest that the Mismatch Asymmetry is driven by the existence of a more general, and hence mismatch-independent, penalty for passive ellipsis clauses, one that affects acceptability regardless of whether the antecedent and ellipsis clause differ in voice. Importantly, the data do not support the existence of a penalty against passive clauses more

generally, as that would predict no difference between the [P -> A] and [A -> P] mismatches, as well as a greater degree of degradation of [P -> P] matches as compared to the other conditions. Instead, the data is explained best by an additive combination of a mismatch penalty and a passive ellipsis clause penalty.

3.2 Experiment 2

The results from Experiment 1 appear to be problematic for the RH, since the hypothesis offers no explanation for why a parallel penalty for passive ellipsis clauses would be witnessed in the matched condition. The results instead support the existence of a more general passive ellipsis clause penalty (henceforth, PECP) – whatever the underlying explanation for it might be – and that the mismatch asymmetry is merely a by-product of that effect.

Whereas Experiment 1 demonstrated that the PECP is specific to passive clauses (that is, as compared to actives), it did not establish that it is specific to ellipsis. Demonstrating this requires that the behavior of non-elliptical controls also be examined, since only when the respective behaviors of elliptical cases and their non-elliptical variants diverge can we conclude that a found effect is attributable to ellipsis *per se*. In Experiment 2, we therefore ask whether we find similar evidence for a passive penalty in discourses that do not contain ellipsis. If so, that would suggest that the explanation for the Mismatch Asymmetry lies outside of the theory of ellipsis. If not—a finding that would be consistent with previous studies of VP-ellipsis that have utilized non-ellipsis controls (Kim et al., 2011; SanPietro et al., 2012; Kim & Runner, 2018)—it would suggest the need for an explanation that is particular to the linguistic properties of ellipsis. The purpose of Experiment 2 is thus two-fold: to replicate the results found for ellipsis clauses in Experiment 1, and to examine whether similar effects occur for variants of the stimuli from which nothing has been elided.

3.2.1 Methods

Stimuli

In addition to the stimuli used in Experiment 1 exemplified in (110) and (111), a no-ellipsis condition was added with variants in which the elided VP was made overt, as in (113). Following Kim and Runner (2018), the overt VP was reduced linguistically as much as possible, e.g. by pronominalizing NPs wherever it was felicitous to do so, in order to mitigate a potential independent penalty associated with producing overt material that could have been elided.

(113) a. The judge read the report first, and then the lawyer read it too. [A -- A]

b. The report was first read by the judge, and then the confession was read too. [P -- P]

c. The report was first read by the judge, and then the lawyer read it too. [P -- A]

d. The judge read the report first, and then the confession was read too. [A -- P]

The result was a set of 24 experimental items following a 2x2x2 design, crossing three independent factors: whether the two clauses were MATCHED or MISMATCHED in voice, whether the two clauses were ACTIVE or PASSIVE, and whether the second clause was ELIDED or UNELIDED.

Procedure

60 self-reported native speakers of English were recruited via Amazon.com's Mechanical Turk. The procedure remained the same as in Experiment 1: in a within-item and within-participant design, each participant was presented with 24 experimental items and 48 filler items, and performed an acceptability judgment task using a 5-point Likert scale. As before, the experiment was conducted using the Ibex software.

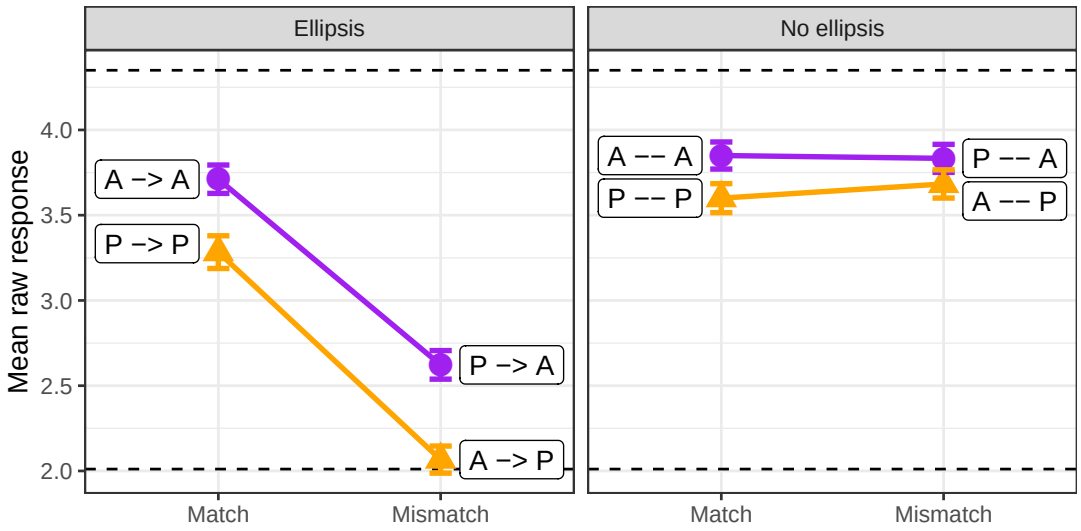


Figure 3.2: Results from Experiment 2. Errorbars reflect Standard Errors, dashed lines indicate mean ratings of (un)acceptable elliptical fillers

3.2.2 Results

We conducted a cumulative probit mixed-effects regression analysis that was analogous to the one for Experiment 1, but was extended to capture the full 2x2x2 design of Experiment 2. We included sum-coded population-level effects for each condition in the experiment and all interactions, as well as the maximal group-level effect structure for items and subjects. The results are summarized in Figure 3.2.

First, there was a significant mismatch penalty overall (i.e., averaging over all other factors; $\Delta = -0.34, CI(\Delta) = [-0.46, 0.22], P(\Delta < 0) = 1$), which was greater under ellipsis ($\Delta = -0.37$, $CI(\Delta) = [-0.5, 0.26], P(\Delta < 0) = 1$). Secondly, there was a significant overall passive penalty ($\Delta = -0.22, CI(\Delta) = [-0.36, -0.08], P(\Delta < 0) = 1$), which was also significantly greater under ellipsis ($\Delta = -0.09, CI(\Delta) = [-0.19, 0], P(\Delta < 0) = 0.97$), but did not interact with mismatch ($\Delta = -0.01$, $CI(\Delta) = [-0.09, 0.08]$, $P(\Delta < 0) = 0.57$). (The 3-way interaction between all conditions was not significant: $\Delta = -0.05$, $CI(\Delta) = [-0.12, 0.03]$, $P(\Delta < 0) = 0.89$.) These results fully replicate the results from Experiment 1 (see left panel in Figure 3.2): under ellipsis, there were two independent main effects—a mismatch penalty and a penalty for passive ellipsis

clauses—and no interaction between them. Going beyond Experiment 1, however, Experiment 2 further suggests that both of these effects are (at least partially) ellipsis-specific, as evidenced by the significant interactions between mismatch/voice and ellipsis.

3.2.3 Discussion

The goal of Experiment 2 was to assess whether the passive penalty found in Experiment 1 is an ellipsis-specific effect. Broadly consistent with the result of previous studies (Kim et al., 2011; SanPietro et al., 2012; Kim & Runner, 2018), the results confirm that it is. Whereas the results in the ellipsis condition revealed the same significant effect for passive ellipsis clauses seen in Experiment 1, there was no analogous significant effect in the no-ellipsis condition, suggesting that the effect is at least partially specific to ellipsis. This result rules out, among other things, the existence of a more general, discourse-based penalty for passive sentences in the types of passages utilized in our experiments.

3.3 Experiment 3

To summarize thus far, we have two competing hypotheses regarding the Mismatch Asymmetry: the PECP hypothesis and the RH. The PECP hypothesis accounts for the fact that `[A -> P]` mismatches are rated as less acceptable than `[P -> A]` mismatches because only the former contain a passive ellipsis clause. On the other hand, the RH posits that, given a particular pairing between an antecedent clause and an ellipsis clause, the structure associated with the antecedent clause is subject to misremembering. Because previous work suggests that passives are misremembered as actives more often than actives as passives, the RH posits that `[P -> A]` mismatches are more likely to yield an "illusion of grammaticality" than `[A -> P]` mismatches. Crucially, this prediction rests on the fact that the antecedent clause comes before the ellipsis clause, and hence is the clause that is subject to misremembering.

The results of Experiments 1 and 2 supported two predictions that are shared by the PECP and RH: the existence of the Mismatch Asymmetry, and the lack of a similar effect in unelided variants of the same discourses. However, only the PECP predicts a third effect that was confirmed in Experiment 1: that `[P -> P]` cases are rated as less acceptable than `[A -> A]` cases. At best, the RH is silent on such effects, since the Recycler is not hypothesized to be engaged in cases of syntactic match. Therefore, on the assumption that the penalty on passive clauses found in both matched and mismatched ellipses is the result of a common cause, the RH misses an important generalization.

In order to more definitively compare the two explanations of the Mismatch Asymmetry, however, a type of example is needed for which the hypotheses make both crisp and opposing predictions. Fortunately, there is such a case. As is well-known, VP-ellipsis is acceptable when used cataphorically in subordinate discourse configurations, as in (114):

(114) If he wants to, the judge will read the report.

Example (114) is acceptable even though the ellipsis clause *he wants to* precedes the catacedent clause *the judge will read the report*. This referential pattern mirrors that of the pronoun in (114), whereby reference with *he* is successful despite preceding its catacedent *the judge*.

The PECP hypothesis and RH make opposite predictions for such cases. On the one hand, the predictions of the PECP hypothesis are as before: mismatches that contain a passive ellipsis clause (and hence an active catacedent clause) should be judged as less acceptable than those involving an active ellipsis clause and a passive catacedent. That is, the ordering of the clauses shouldn't matter. The RH, on the other hand, makes the opposite prediction: mismatches that contain a passive ellipsis clause and active catacedent clause should be judged as *more* acceptable than those that contain an active ellipsis clause and a passive catacedent. This prediction results from the fact that, by virtue of being the initial clause, it is the *ellipsis* clause that is subject to misremembering. That is, upon encountering a cataphoric ellipsis site, the processor will

anticipate, and ultimately identify, the occurrence of the catacedent. When the processor attempts to establish IDENTITY between the catacedent and the (invisible) structure at the ellipsis site, it is the ellipsis clause that has to be retrieved from memory. As illustrated in (115), since `[P <- A]` mismatches require the retrieval of a passive clause from memory, they are more likely to elicit an illusion of grammaticality than `[A <- P]` mismatches, which involve the retrieval of an active clause.

(115) a. Before the lawyer did, the report was first read by the judge. `[A <- P]`

b. *Before the confession was*, the judge read the report first. illusory `[A <- A]`

The purpose of Experiment 3 is to evaluate these competing predictions.[10]

3.3.1 Methods

Stimuli

Cataphoric variants of the stimuli used in Experiment 1 were constructed. Recall that half the stimuli from Experiment 1 employed *after* to connect the clauses, as in (111). Since *after* is a subordinating conjunction, the variants could be constructed simply by reversing the order of the clauses, as in (116).

[10] A reviewer questions whether the predictions of the RH for cataphora are as straightforward as our characterization would suggest, noting that it is possible that cataphoric and non-cataphoric cases are processed in different ways. Specifically, the reviewer suggests that the forward-looking dependency created by cataphoric VP-ellipsis might lead to stronger maintenance of the initial clause in active memory when the matrix clause is processed. An increase in the memory trace of this sort would in turn predict that mismatches that contain a passive ellipsis clause and an active catacedent should be as (un)acceptable as those with an active ellipsis clause and a passive catacedent, since cataphoric passive ellipsis clauses would not be subject to the misremembering effect.

We are admittedly not completely clear on the logic underlying this suggestion (for instance, why cataphora would lead to greater attention on the *initial* clause rather than the *final* clause, the latter of which will ultimately resolve the dependency), and believe that additional evidence of such a processing difference would be required for this proposal to have sufficient argumentative force. But even if we grant the possibility the reviewer outlines, Experiment 3 still provides an adequate test of the RH. Specifically, if the first clause does receive a memory boost due to the cataphoric VP-ellipsis, it should do so for both passive and active ellipsis clauses alike, in turn predicting the elimination of a mismatch asymmetry for cataphora. The results will instead show that the effect persists, in the manner captured by the PECP.

(116) a. After the critic did already, the customer praised the dessert. [A <- A]

b. After the appetizer was already, the dessert was praised by the customer. [P <- P]

c. After the critic did already, the dessert was praised by the customer. [A <- P]

d. After the appetizer was already, the customer praised the dessert. [P <- A]

On the other hand, the other half of the stimuli used in Experiment 1 employed the coordinating conjunction *and*, and hence the second clause could not be fronted to form a subordinate structure. We therefore adapted the examples to employ the subordinate connective *before*, as in (117).

(117) a. Before the lawyer did, the judge read the report first. [A <- A]

b. Before the confession was, the report was first read by the judge. [P <- P]

c. Before the lawyer did, the report was first read by the judge. [A <- P]

d. Before the confession was, the judge read the report first. [P <- A]

This yielded a set of 24 stimuli, 12 utilizing *after* as the connective, and 12 utilizing *before*.

Procedure

As in Experiment 1, we recruited 30 participants from Amazon.com's Mechanical Turk and presented with 24 experimental stimuli alongside 48 filler items in an acceptability judgment task via Ibex. Two participants reported being non-native English speakers and were therefore excluded from all analyses. The details of the design and task followed those given in Section 3.1.1.

3.3.2 Predictions

Two predictions carry over from Experiment 1. First, we expect an effect of mismatch, whereby [A <- P] and [P <- A] cases are judged to be less acceptable than [A <- A] and [P <- P] cases. Second, the PECP hypothesis, but not the RH, predicts that [P <- P] examples

will be judged as less acceptable than the [A <- A] examples. On the hypothesis that the PECP is an independent factor that co-exists with the mismatch penalty, the PECP hypothesis does not predict an interaction.

The key prediction concerns the relative level of acceptability of [A <- P] and [P <- A] mismatches. As explained above, the PECP hypothesis predicts that [P <- A] mismatches will be judged as less acceptable than [A <- P] mismatches, since the former contains a passive ellipsis clause. The RH, on the other hand, predicts that [P <- A] mismatches will be judged as more acceptable than [A <- P] mismatches, since passive initial clauses are more easily misremembered as active clauses than active initial clauses remembered as passive.

3.3.3 Results

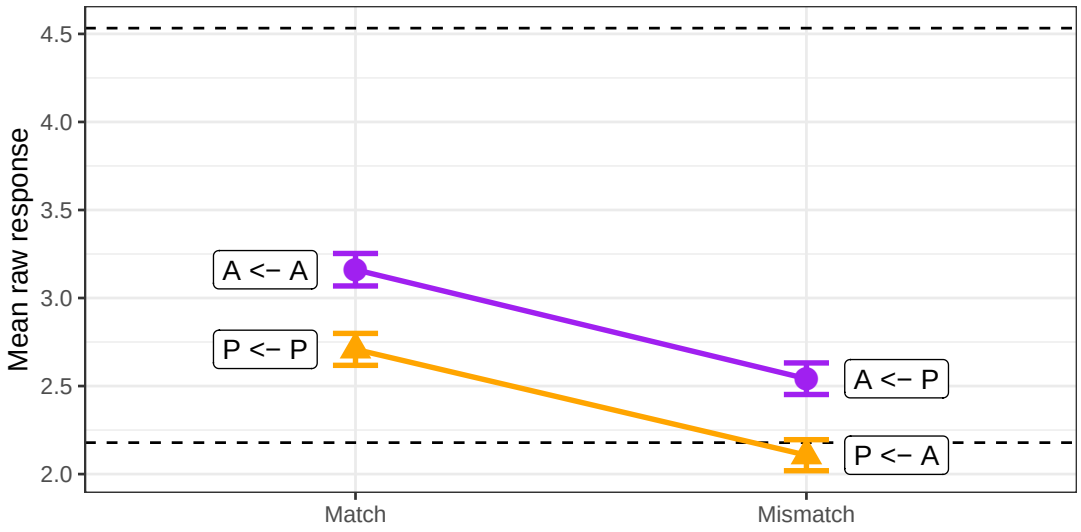


Figure 3.3: Results from Experiment 3. Errorbars indicate Standard Errors, and dashed lines show mean ratings of (un)acceptable elliptical fillers

As in Experiments 1 and 2, we analyzed raw acceptability in a cumulative probit mixed-effects regression with MISMATCH and VOICE of the ellipsis clause and the interaction between the two as population-level effects and the all group-level effects permitted by the design. The results from Experiment 3 are summarized in Figure 3.3. As in Experiment 1, there were two significant

main effects: a mismatch penalty ($\Delta = -0.73$, $CI(\Delta) = [-1.04, -0.44]$, $P(\Delta > 0) = 1$), and a penalty for passive ellipsis clauses ($\Delta = -0.58$, $CI(\Delta) = [-1.04, -0.44]$, $P(\Delta > 0) = 1$). There was no evidence that the passive penalty differed across matched/mismatched conditions ($\Delta = -0.07$, $CI(\Delta) = [-0.49, 0.35]$, $P(\Delta > 0) = 0.62$).

3.3.4 Discussion

The results confirm the predictions of the PECP hypothesis and run counter to those of the RH. First, as in Experiment 1, the penalty on acceptability for passive ellipsis clauses was not limited to the mismatch condition; instead, there was an analogous difference in the matched condition. This result is consistent with the PECP hypothesis, but cannot be explained by the RH since there is no grammatical violation to repair in the matched cases. Second, the manipulation of clause order did not have the effect predicted by the RH: `[P <- A]` mismatches were judged as less acceptable than `[A <- P]` mismatches. This result is consistent with the PECP hypothesis, but should have gone in the opposite direction according to the RH, since passage-initial, passive ellipsis clauses should be more likely to be misremembered as active than active ellipsis clauses misremembered as passive.

3.4 General discussion

We set out to explore the source of the Mismatch Asymmetry, in part by evaluating the predictions of the RH account on previously unexamined cases. Experiment 1 sought to replicate the asymmetry using stimuli adapted from the original Arregui et al. (2006) study, but using a fuller paradigm that contained voice-matched control items. Whereas the results replicated the key finding of Arregui et al.'s study, they also revealed a penalty for passive ellipsis clauses even when the ellipsis clause is syntactically matched with the antecedent clause. This finding instead suggested the existence of a passive ellipsis clause penalty, or PECP. Experiment 2 then

asked whether the passive penalty might be attributable to a more general, ellipsis-independent condition by also examining unelided variants. Whereas the effect found in Experiment 1 was replicated for the elided versions, no such effect was found for the unelided variants, indicating that the passive penalty is specific to ellipsis clauses. Experiment 3 then provided a critical test of the RH analysis by examining cases that feature cataphoric VP-ellipsis. The results revealed that, like the previous two experiments, mismatches with passive ellipsis clauses are rated as worse than those with active ellipsis clauses. This result runs counter to the RH, as it predicts the opposite effect. Together, therefore, the experiments point to the existence of a PECP that applies across the board to both matched and mismatched cases of VP-ellipsis, in both anaphoric and cataphoric discourse configurations.

In addition to offering a refutation of the RH's explanation of the Mismatch Asymmetry, the results of our study call into question the relevance of the asymmetry to the fundamental question set out in the introduction, specifically regarding at what level(s) of language processing constraints on acceptable usage apply and interpretation mechanisms operate. For one, recall that if the predictions of the RH were confirmed, it would have potentially provided strong support for syntactic analyses, since the misremembering phenomenon to which the RH appeals applies specifically to syntactic representations. That is, no similar explanatory path would appear to be available to proponents of referential theories. Our findings cast significant doubt on the efficacy of the analysis, however, with the result being that this line of argumentation in favor of syntactic accounts is rendered moot. To be clear, the results presented here do not argue *against* syntactic analyses either. Instead, the Mismatch Asymmetry remains a mystery on both syntactic and referential analyses.

A second finding of note is the degraded acceptability of syntactically-matched passive voice ellipses as compared to matched active cases, as found in previous studies and Experiment 1. This result is also surprising for both types of account. It is mysterious on syntactic analyses, since for both `[A -> A]` and `[P -> P]` ellipses, there exists a syntactically-matching, and hence

perfectly suitable, VP available in the syntactic representation of the antecedent clause. As such, there is no constraint violation involved, and hence no need for a recovery mechanism such as the RH. The finding is likewise mysterious for referential accounts, for analogous reasons. That is, in each scenario a suitable representation of the referent has been computed as part of the compositional semantic analysis of the antecedent clause, and hence should be readily available as a referent for a subsequent VP-ellipsis. In neither case are any additional inferential steps needed to fashion an appropriate representation of the referent, as we saw is necessary in cases of syntactic mismatch. The results therefore suggest that there is a penalty for passive ellipsis clauses, one that demands an explanation regardless of which type of account of VP-ellipsis interpretation one adopts.

This raises the obvious question of what the underlying source of the penalty is, such that it is independent of mismatch yet only applies in the context of ellipsis. Whereas we are only in a position to speculate at this time, we suspect that the explanation lies in the domain of information structure. In particular, we hypothesize that the penalty may result from a clash between the respective information structural properties of the passive and of VP-ellipsis, particularly as they relate to the topicality of constituents.

On the one hand, it is well-known that active voice and passive voice constructions differ with respect to their information structural properties. Whereas the active voice construction in English is relatively unmarked with respect to information structure (with a relatively weak tendency for subjects to be construed as topics), one of the primary functions of the passive is to mark its subject as being topical (Shibatani, 1985; Givón, 1990; Rohde & Kehler, 2014, among others). As such, whereas the meaning of any constituent could potentially be topical in (118a), there is a much stronger presumption that the report is topical in (118b).

(118) a. The judge read the report.

b. The report was read by the judge.

Otherwise, it is unclear what a speaker's motivation would be to choose the passive over the unmarked active. Another way to cast the observation is in terms of Question-Under-Discussion (QUD) models of discourse coherence (Roberts, 1998, 2012), according to which topical elements of an utterance are those meanings that are provided by the operative QUD. In this regard, we note that sentence (118a) could serve as a felicitous answer to a variety of implicit questions, e.g., *What happened?* or *What did the judge do?*. Sentence (118b), on the other hand, comes across as a better answer to the question *What happened to the report?*, in which the report is part of the topic. It would be a less natural answer to the question *What did the judge do?*, for instance.

On the other hand, according to referential theories, VP-ellipsis is predicted to behave like other pro-forms, such as entity-referring personal pronouns (see Chapter 2 for extensive discussion). According to some theories of pronoun usage (Gundel et al., 1993; Grosz, Joshi, & Weinstein, 1995; Rohde & Kehler, 2014, among others), pronouns serve an information structural function as well, specifically to indicate a continuation of an entity-level topic. For this reason, whereas the pronoun *He* in passage (119) is to some degree ambiguous between John and Bill, the pronoun *He* in (120), where the choice to use the passive has placed Bill in a strongly topical position, is more likely to be understood to refer to Bill than to John.

(119) John reprimanded Bill. He was upset.

(120) Bill was reprimanded by John. He was upset.

On this logic, if VP-ellipsis involves a pro-form, we expect it to likewise carry a presupposition that its meaning is topical. And this is indeed the case in the stimuli used here and in the previous studies surveyed. Viewing the question again through the lens of QUD analyses, passages like (121) cohere by virtue of their clauses each providing a partial answer to a common QUD, in this case, *What was read by the judge?*.

(121) The report was read by the judge, and the confession was too.

The meaning of the elided VP is presupposed by the operative QUD, it is therefore topical in discourses such as (121).

As a result, conflicting demands occur in a passivized ellipsis clause such as that in (121): the VP-ellipsis requires that the VP meaning (*was read by the judge*) be topical, whereas the speaker has used a construction that indicates that the remnant subject NP (*the confession*) is topical. Clearly one cannot have it both ways, for focus must be present somewhere in the clause. In such a situation, the speaker therefore has at least two other options, both of which are preferable: use the active voice construction, for which the elided VP meaning can felicitously serve as the topic, or use the passive without employing ellipsis, so that the surface subject can serve as the topic without conflicting constraints. This hypothesis thus gives us an explanation for why we see a PECP in both matched and mismatched conditions, but only when VP-ellipsis occurs.

As noted earlier, this hypothesis is only speculative, and it is not among our goals to offer a vigorous defense of it here. However, we do point out that it does make an immediate prediction: that there should be no penalty for passive ellipsis clauses in which the subject and VP meaning can be both construed as being part of the topic. There is in fact evidence to support this prediction. Kertz (2013) conducted an experiment to test her hypothesis that the penalty for syntactic mismatch would vary according to information structural properties associated with ellipsis clauses: in particular, that ellipsis clauses that display AUXILIARY FOCUS will be more resilient to mismatch than those that display SUBJECT FOCUS. Her Experiment 3 employed stimuli of the sort shown in (122a)–(122d):[11]

(122) a. The technicians didn't install the line as quickly as the engineers did.

[subject focus, match]

b. The line wasn't installed by the technicians as quickly as it could have been.

[11]There were also two non-ellipsis variants included in each stimulus set, which we omit here for simplicity.

[auxiliary focus, match]

c. The line wasn't installed by the technicians as quickly as the engineers did.

[subject focus, mismatch]

d. The technicians didn't install the line as quickly as it could have been.

[auxiliary focus, mismatch]

Kertz compared the relative acceptability of cases in which accent falls on the subject of the ellipsis clause, as in (122a) and (122c), with cases in which accent falls on the auxiliary, as in (122b) and (122d). The results revealed a reliable interaction between mismatch and focus, whereby simple focus ellipses were rated as significantly more acceptable in the mismatch condition, but not in the match condition.

Two of Kertz findings provide preliminary support for our hypothesis. First, in contrast to the results of our Experiment 1, there was no passive ellipsis clause penalty witnessed in the matched condition: `[P -> P]` ellipses such as (122b) were not reliably rated as less acceptable than `[A -> A]` ellipses such as (122a). This is explained by the fact that her `[P -> P]` stimuli featured ellipsis clauses in which focus resided only on the auxiliary. As such, both the passivized subject and the VP meaning are topical, and no PECP resulted. Second, whereas her results revealed a mismatch asymmetry, it went in the opposite direction as the one found here and in other previous work: `[A -> P]` mismatches such as (122d) were rated as *more* acceptable than `[P -> A]` mismatches such as (122c). Again, this is consistent with the hypothesis, since her `[A -> P]` mismatches, unlike her `[P -> A]` mismatches, were cases of auxiliary focus, and hence the other elements of the sentence, including the meanings of both the subject NP and the elided VP, were topical. This suggests that there was no PECP in effect to bring down the ratings of the `[A -> P]` mismatches.

Therefore, as an ensemble, the foregoing evidence suggests that there is a PECP, but one that applies only in ellipsis clauses that bear subject focus. These are just the cases in which

the need for the meanings of both the subject and the elided NP to be topical are in conflict: the elision of the VP in turn requires that focus falls on the subject, whereas one of the central functions of the passive is to mark its subject as topical. This conflict does not exist in auxiliary focus constructions, and hence we find no evidence of a penalty.[12]

A reviewer of the manuscript published in Poppels and Kehler (2019) rightfully asked whether the PECP is unique to VP-ellipsis, or if it extends to other types of ellipsis more generally. To gain insight into this question, we carried out a pilot experiment to investigate the potential existence of a PECP in two other forms of ellipsis, specifically gapping and sluicing. 24 items were derived from the items used in Experiment 1, with 12 containing gapping as in (123) and 12 containing sluicing as in (124).

(123) a. Mary scolded Wilma, and Susan, Nancy.

b. Wilma was scolded by Mary, and Nancy, by Susan.

c. Mary scolded Wilma, and Susan scolded Nancy.

d. Wilma was scolded by Mary, and Nancy was scolded by Susan.

(124) a. Someone read the report, but I don't know who.

b. The report was read by someone, but I don't know by whom.

c. Someone read the report, but I don't know who read it.

d. The report was read by someone, but I don't know by whom it was read.

Each item followed a 2x2 design that crossed VOICE (active vs. passive) with ELLIPSIS (ellipsis

[12]It is therefore perhaps unsurprising that attested cases of `[A -> P]` mismatches cited in the literature, such as (i)–(ii) from Kehler (2002b), are characterized by auxiliary focus rather than subject focus:

(i) Actually I have implemented it [= a computer system] with a manager, but it doesn't have to be. [implemented with a manager] (Steven Ketchpel, in conversation)

(ii) Just to set the record straight, Steve asked me to send the set by courier through my company insured, and it was. [sent by courier through my company insured] (posting on the Internet)

vs. no ellipsis).[13]

26 native speakers of English, recruited via Amazon.com's Mechanical Turk, participated in the pilot and the results are shown in Figure 3.4. As for the previously reported experiments, we used the brms R package to test for the presence of main effects of voice and ellipsis, considering gapping and sluicing items separately. Since there were only 12 items for each ellipsis type instead of 24, the pilot study has substantially lower statistical power than the main experiments. Consequently, the lme4 analysis reported in Poppels and Kehler (2019) suffered from convergence issues, which forced us to reduce the random effects structure (following the recommendations in Barr et al., 2013) in order to achieve convergence. Using the brms package, however, model convergence is largely a function of collecting a sufficient number of posterior samples, and as a result we were able to fit maximal hierarchical models with item- and subject-specific intercepts and slopes corresponding to all hypotheses. None of the results, which are reported below, are qualitatively different from the ones reported in Poppels and Kehler (2019).

For sluicing (right side of Figure 3.4), whereas there was a numerical difference between active and passive item variants (vertical distance between lines in the graph), the statistical analysis revealed that it was not significant ($\Delta = -0.19$, $CI(\Delta) = [-0.79, 0.4]$, $P(\Delta < 0) = 0.75$). There was, however, a significant difference between elliptical and non-elliptical item variants in that items involving sluicing significantly improved in acceptability compared to their non-

[13]The design did not include cases that involve syntactic mismatch, in part because it is not clear how to do so for sluicing. Specifically, whereas it is possible to construct cases that unambiguously involve an active-passive mismatch as in (i),

(i) Someone read the report, but I don't know who by. [the report was read]

we see no way to construct cases that force a passive-active mismatch as in (iia),

(ii) a. The report was read by someone, but I don't know who. [read the report]
b. The report was read by someone, but I don't know who. [the report was read by]

since an alternative analysis as a passive-passive match will also be available, as in (iib).

Although it is possible to test mismatched versions of gapping constructions, we opted to keep the designs for sluicing and gapping invariant.

elliptical counterparts ($\Delta = 0.31, CI(\Delta) = [0.08, 0.54], P(\Delta > 0) = 0.99$). The interaction between ellipsis and voice was not significant ($\Delta = 0.06, CI(\Delta) = [-0.16, 0.29], P(\Delta < 0) = 0.28$).

For gapping (left side of Figure 3.4), ellipsis had a large, negative effect on acceptability ($\Delta = -0.99, CI(\Delta) = [-1.48, -0.53], P(\Delta < 0) = 1$). There was no main effect of voice ($\Delta = 0.2, CI(\Delta) = [-0.38, 0.79], P(\Delta < 0) = 0.24$). The interaction between voice and ellipsis reached marginal significance ($\Delta = 0.3, CI(\Delta) = [-0.12, 0.74], P(\Delta > 0) = 0.92$), whereby items involving passive gapping tended to be more acceptable than those involving active gapping, but with no such difference between their respective non-elliptical counterparts.

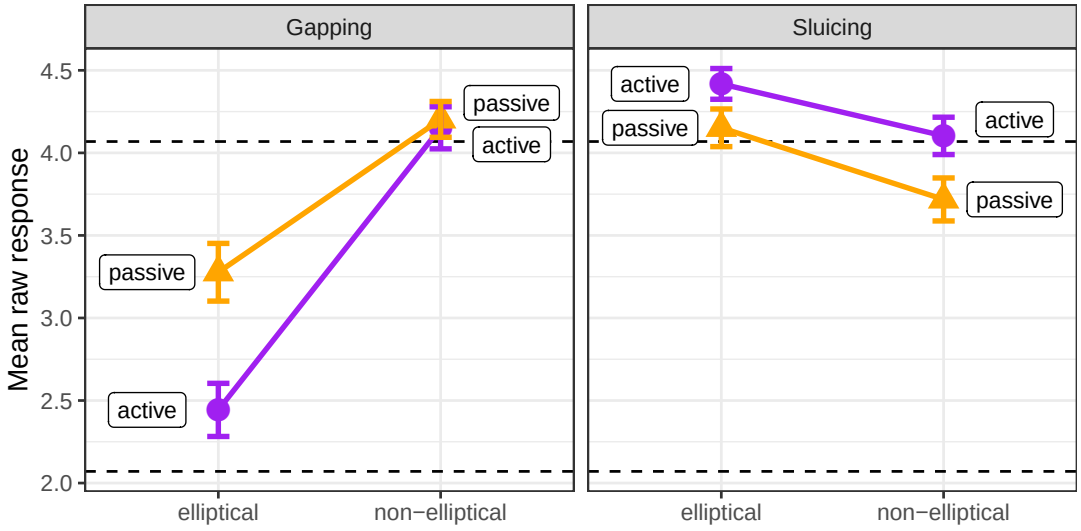


Figure 3.4: Results from pilot study for gapping (left) and sluicing (right). Errorbars reflect Standard Errors, dashed lines indicate mean ratings of (un)acceptable elliptical fillers

Whereas this study was merely a pilot, two provisional conclusions nonetheless emerge. First, the PECP that we found for VP-ellipsis is not operative in either of these other forms of ellipsis. Whereas the sluicing data show a numerical trend toward a passive clause penalty – one which might or might not become significant in a full experiment – any such penalty appears to apply equally to the non-elliptical variants, unlike what was found for VP-ellipsis in Experiment 2. On the other hand, whereas a difference was found between ellipsis and no-ellipsis clauses in the gapping data, passive ellipsis clauses were actually rated more highly than active ones.

Second, it is clear that the different forms of ellipsis display varied patterns of behavior, ones that may ultimately be tied to information structural factors that are specific to the linguistic properties of the respective forms. For instance, the robust penalty against no-ellipsis clauses found in the sluicing condition is likely due to a "repeated-clause" penalty, associated with the overt expression of a clause that could have been felicitously sluiced due to a representation of it being readily recoverable from the context. On the other hand, for gapping we found that ellipsis had a large, negative effect on acceptability. This negative effect might result from the fact that gapping sentences tend not to occur "out-of-the-blue," but instead only when an "open proposition" (or equivalently, a multi-Wh QUD) exists in the context or can be accommodated from it (Sag, 1976; Wilson & Sperber, 1979; Prince, 1986; Steedman, 1990).[14] As such, the effect may plausibly owe to the fact that the experimental stimuli failed to provide the sort of contexts necessary for gapping to be fully felicitous.

We have likewise offered an information structural explanation of the PECP, one that, of the different forms of ellipsis examined here, would only be expected to apply in the case of subject-focus VP-ellipsis for the reasons previously noted. Clearly, however, future work will be necessary to fully explore the mechanics and resulting predictions of this hypothesis; our goal here is merely to offer it as a possible line of investigation. Our primary goal instead has been to demonstrate the inadequacy of memory-based explanations of the Mismatch Asymmetry such as the RH. As we have seen, this type of explanation fails to capture the generalization that the penalty for passive ellipsis clauses applies equally to both matched and mismatched ellipses (Experiment 1), the fact that the dispreference for passive ellipsis clauses remains when VP-ellipsis is used cataphorically (Experiment 3), and the reversal of judgments that Kertz found when focus structure is manipulated.

[14]For example, Steedman (1990, p. 248) says: "Indeed, even the most basic gapped sentence, like *Fred ate bread, and Harry, bananas*, is only really felicitous in contexts which support (or can accommodate) the presupposition that the topic under discussion is *Who ate what*."

3.5 Conclusion

Previous work has revealed the existence of a Mismatch Asymmetry, whereby cases of mismatched VP-ellipsis with passive ellipsis clauses and active antecedent clauses are regarded as less acceptable than cases with active ellipsis clauses and passive antecedents. The most prominent explanation of the asymmetry is that provided by the RH: because passive clauses are more prone to be misremembered as active than the other way around, mismatches that involve a passive antecedent clause and active ellipsis clause are more likely to yield an "illusion of grammaticality" than cases that involve an active antecedent with a passive ellipsis clause. The RH not only stands to explain the effect, but potentially has broader ramifications as well. Specifically, because only syntactic, and not semantic, representations are prone to misremembering, a demonstration of the correctness of the RH proposal would also provide significant support for a syntactic analysis of VP-ellipsis over a referential one.

We have provided the results of three experiments that explored the source of the Mismatch Asymmetry, with particular attention to the predictions of the RH. Experiment 1 replicated the asymmetry, but also revealed that a parallel penalty occurs for syntactically matched cases in the passive voice as well. The results therefore suggest that the source of the penalty is more general than the domain over which the RH applies. Experiment 2 examined whether similar penalties are witnessed in variants in which there is no ellipsis. Consistent with the RH, the penalty did not generalize to the unelided variants. Experiment 3 provided a critical test of the theory by employing variants in which VP-ellipsis refers cataphorically. Whereas the RH predicts that cases with active voice ellipsis clauses should be rated as less acceptable than those with passive voice ellipsis clauses, the results revealed the opposite effect. In total, the results of Experiments 1-3 reveal consistent evidence for a penalty against ellipsis clauses in the passive voice. We therefore conclude that the explanation offered by the RH fails to explain the data, and hence that it provides no evidence, either for or against, syntactic theories of VP-ellipsis.

What remains is the question of what the underlying source of the PECP is. We have speculated that the cause is ultimately information-structural, bearing particularly on a conflict in the placement of topic and focus within passive ellipsis sentences. This hypothesis is consistent with our data as well as that of Kertz (2013), who demonstrated that the Mismatch Asymmetry can be reversed by manipulating the focus structure of the ellipsis clause. Further research will be required to uncover the ultimate explanation of the phenomenon.

3.6 Note regarding previous publication

The material presented in this chapter was co-authored by Andy Kehler and has been published in *Glossa: A Journal of General Linguistics*:

> Poppels, T., & Kehler, A. (2019). Reconsidering asymmetries in voice-mismatched verb phrase ellipsis. *Glossa: A Journal of General Linguistics*, *4*, 1–22. doi:http://doi.org/10.5334/gjgl.738.

It was adapted for this thesis by (i) removing material from the introduction that was rendered redundant in the context of Chapter 2, and (ii) by reporting improved statistical analyses of all experimental results compared to the ones reported in Poppels and Kehler (2019).

Chapter 4

Inferential VP-ellipsis

It is well established that Verb Phrase Ellipsis (VPE), illustrated in (125), canonically exhibits a close correspondence between the ellipsis clause and the antecedent clause.

(125) The students had to study hard because the teacher didn't (study hard).

This correspondence manifests itself in two ways (see Chapter 2 for more details). First, the acceptability of VPE appears to depend on the existence of some form of parallelism between the ellipsis clause and the antecedent clause. Second, when such parallelism constraints are met, the set of possible interpretations of the ellipsis clause tends to be highly restricted by the meaning of the antecedent clause. In the case of (125), for example, the meaning of the ellipsis clause must involve the meaning of the antecedent VP *study hard*; the ellipsis simply cannot be resolved to the proposition that the teachers didn't teach the material well, for instance, even though that would arguably result in a more plausible interpretation.

As I detailed in Chapter 2, there are two ways in which these constraints have been captured in theories of VPE. The first has been to posit the existence of particular constraints on the felicitous use of VPE, ones that impose a requirement for IDENTITY between the antecedent and ellipsis clauses at some level of linguistic representation (most notably, at the level of syntax

or semantics). The second is to treat VPE as a form of discourse reference, thereby assimilating its behavior to that of other familiar forms of reference, such as pronouns. Associated with each type of theory is a set of well-known problems. IDENTITY theories have to contend with cases in which VPE is acceptable and interpretable despite the lack of a suitable antecedent, which has motivated researchers to propose various rescue strategies to preserve IDENTITY. Referential theories, on the other hand, need to explain why various cases of VPE are infelicitous despite the fact that the required referent could plausibly be seen to be inferrable from the discourse context.

We begin by reviewing both theories and the predictions they make with respect to certain cases of VPE. We then present results from an acceptability judgment experiment using a type of example not previously discussed in the literature, which, we argue, are only consistent with a referential account. Although the results ultimately lead us to advocate for the referential account, we conclude by discussing some of the remaining challenges for such theories.

4.1 IDENTITY theories of VPE

As noted above, the correspondence illustrated with (125) has inspired a class of IDENTITY theories of ellipsis, which posit that eliding linguistic material is only grammatical when it is identical (in some relevant way) to an antecedent VP (Hankamer & Sag, 1976; Sag, 1976; Chung et al., 1995; Merchant, 2001; Chung, 2013; Rudin, 2019, and many others). Much of the previous literature on VPE has focused on arbitrating among different types of IDENTITY theories or determining at which level of representation IDENTITY is evaluated—a project that Lipták (2015) referred to as the "quest for identity," and which I described in detail in Section 2.1. Here we collapse different flavors of IDENTITY theories in order to focus on an important property they have in common, specifically that they reduce the constraints on VPE to the linguistic antecedent, without appealing to non-linguistic aspects of the context.

Recall from Chapter 2 that the attempt to reduce the context-dependency of VPE to

an IDENTITY relation between the ellipsis site and the linguistic antecedent makes IDENTITY theories vulnerable to a variety of counterexamples that are acceptable and interpretable but lack the kind of antecedent VP deemed necessary. Consider (126), for instance.

(126) a. **Lexical mismatches:** John didn't see anyone, but Mary did ~~see someone~~.
(Merchant, 2013a)

b. **Split antecedents:** John wanted to go to Bolivia and Mary wanted to go to Peru, but because it's expensive, neither of them can ~~go to Bolivia or Peru, respectively~~.
(adapted from Webber, 1978)

c. **Exophora:** [Situational context: Two kids standing in front of a pool]
First kid: I will ~~jump~~ if you will ~~jump~~. (Chao, 1988, cited in Hardt, 1993)

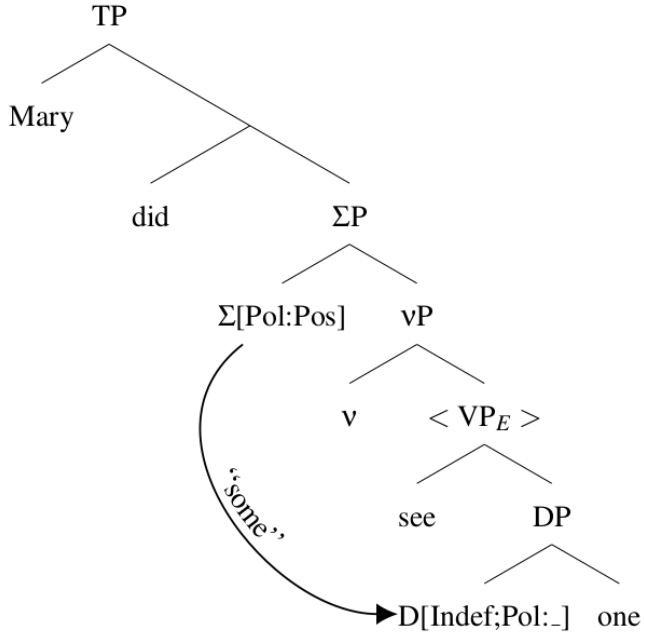
On a syntactic IDENTITY theory, the VP *see anyone* in example (126a) would be ungrammatical at the ellipsis site (**Mary did see anyone*). Instead, the VP required to derive the attested meaning would be *see someone*, but eliding it should be ungrammatical for lack of an identical antecedent. Similarly, in example (126b), there is no single VP that can serve as the antecedent; the necessary VP is instead 'split' across two distinct clauses. Finally, example (126c) involves exophora, in which the meaning of the missing VP is recovered from the situational, and hence non-linguistic, context. Thus, all three examples in (126) showcase perfectly acceptable and interpretable VPE despite the fact that neither of them provide an antecedent VP that is identical to the elided material.

Various proposals have been offered to deal with cases in which VPE is acceptable despite the violation of IDENTITY, which can be roughly classified into two categories: REPRESENTATIONAL approaches, which re-analyze mismatching linguistic elements in a way that preserves IDENTITY, and INFERRED-ANTECEDENT approaches, which invoke cognitive machinery to infer an identical antecedent if the linguistic context fails to provide one.

Representational approaches involve re-analyzing elements that appear non-identical

across the antecedent and ellipsis sites.[1] For example, syntactic IDENTITY can be preserved VP-internally if we analyze non-identical elements in terms of mismatching values that are assigned post-syntactically to otherwise identical abstract elements by agreement with a VP-external phrasal head. To illustrate, consider the following analysis of the lexical mismatch involving polarity items, repeated from (35), which is due to Merchant (2013a):

(127) (see Merchant, 2013a, ex. 3)



The polarity item in the elided VP *see anyone* (as well as the one in the antecedent not shown here) is analyzed as an abstract lexical item at the level of syntax, which is lexicalized post-syntactically as either *some* (as in this example) or *any*, in agreement with the head of the VP-external polarity phrase. The apparent lexical mismatch between *any* in the elided VP and *some* in the antecedent is thus moved outside the elided constituent and attributed to the post-syntactic lexicalization process. This analysis thus establishes the isomorphism between antecedent and elided VP that syntactic IDENTITY theories require. Although we illustrate the logic behind the representational approach to non-identity with this specific example, similar strategies have been applied to a

[1] See Section 2.1.2 for additional examples of this strategy in the context of category mismatches involving nominal antecedents.

variety of other types of mismatches as well.[2]

Another strategy for reconciling IDENTITY theories with cases of non-identity involves postulating cognitive machinery that is capable of inferring a suitable antecedent when the context does not provide one. One example of this strategy is known as ANTECEDENT ACCOMMODATION (Fox, 1999; Thoms, 2015), which extends the formulation of IDENTITY to include a set of "accommodatable" antecedents in addition to the one the context provides.[3] For example, Thoms (2015, ex. 51) posits the following mechanism for accommodating alternative antecedents, adapting similar machinery proposed by Katzir (2007) for generating alternative utterances to be used for deriving implicatures (see also Section 2.1.1):

(128) **Accommodating alternative antecedents for ellipsis**

a. A set of additional antecedents, Ad(A), may be accommodated on the basis of the original (overt) antecedent A.

b. The members of Ad(A) are alternatives derived from A by

   (i) deletion

   (ii) contraction

   (iii) substitution

c. Complexity constraint: all members of Ad(A) must be at most as complex as the overt antecedent A.

d. Semantic constraint: all members of Ad(A) must be semantically identical to the overt antecedent A.

Extending the IDENTITY requirement to the set of accommodatable antecedents Ad(A) makes it possible to capture certain mismatches, such as the polarity-item mismatch described above,

[2]For instance, see Merchant (2013b) for an application of this analytic strategy to voice mismatches.

[3]Another influential proposal for inferring antecedents is the RECYCLING hypothesis (Arregui et al., 2006; Frazier, 2013), which is discussed in depth in Section 2.1.3 and Chapter 3. While we focus our discussion in this chapter on antecedent accommodation, all arguments extend straightforwardly to Recycling-based analyses.

so long as the mismatching elements can be replaced with ones that preserve IDENTITY. Since this is a powerful mechanism that, if left unrestricted, would be prone to overgeneration, Thoms (2015) stipulates additional constraints that require accommodated antecedents to be semantically identical to and at most as complex as the overtly provided antecedent. It is important to note at this point that this notion of accommodation is independent of discourse-level presupposition accommodation in the Lewisian sense (Lewis, 1979). Antecedent accommodation operates on syntactic representations by deleting, contracting, or substituting words and phrases, whereas presupposition accommodation operates on discourse commitments and the common ground between interlocutors.

4.2 Overcoming the IDENTITY crisis: a referential theory of VPE

On a referential account (Webber, 1978; Hardt, 1992, 1993; Kehler, 1993a, 1993b; Hardt, 1999; Kehler, 2000), VPE is not associated with a set of ellipsis-specific rules that impose constraints on identity, but instead operates as a null pro-form. Hence, VPE interpretation is predicted to behave in a way similar to other pro-forms in the language. Crucial to contemporary theories of reference is the observation that reference cannot be aptly characterized by way of only a relationship among pieces of associated linguistic material, i.e. between a pro-form and its antecedent. Instead, the felicitous use of referring expressions depends on the status of the referent with respect to the hearer's knowledge state, which includes (but is not limited to) their evolving mental model of the discourse. This mental model is referred to as the DISCOURSE MODEL, which contains a structured record of the entities and eventualities that have been introduced and the relationships that hold among them (Karttunen, 1976; Webber, 1978, *inter alia*, see Section 2.2 for more details). The felicity of a reference might thus depend on factors such as whether the speaker believes that the hearer has prior knowledge of a referent, whether the

referent has been mentioned previously in the discourse, whether the referent is situated in the immediate surroundings, how salient the speaker thinks the referent is in the hearer's mind, the level of topicality of the referent, and the degree of difficulty in accommodating a discourse model representation for a referent when one doesn't already exist.

On the referential theory, therefore, VPE succeeds by virtue of a suitable representation of the referent existing in, or being constructable from, the hearer's discourse model. Consider (125) again, repeated in (129), and augmented with the representation for the pro-form ϕ:

(129) The students had to study hard because the teacher didn't ϕ. [study hard]

Among the discourse model representations generated as a result of the compositional semantic analysis of the sentence (which will also include representations for the meanings of noun phrases such as *the students* and *the teacher*) will be a representation for the meaning of the VP *study hard*. This discourse model representation thus provides a suitable referent for the subsequent VPE. The discourse model does not, however, contain the VP meaning necessary for deriving the more plausible proposition *the teacher didn't teach the students well*, which is therefore not a possible interpretation of the ellipsis clause.

Support for a referential theory of VPE comes from the fact that it patterns with other types of discourse reference with respect to series of diagnostic properties, which are outlined in detail in Section 2.2.2. For our present purpose, the most relevant property is the possibility of inferential reference resolution: under certain conditions an anaphoric pronoun can access a referent that is not introduced directly by, but is nonetheless inferrable from, an antecedent expression. Consider (130).

(130) Jean is a Frenchman, although he hasn't lived there for many years. (Ward et al., 1991)

Here, the pronoun *there* successfully refers to France, which has not been explicitly mentioned.

However, the referent has a transparent semantic relationship to the meaning of the antecedent *Frenchman*, and hence is derivable from it. Similarly, Webber (1978) offers an example of a related sort involving VPE:

(131) Martha and Irv wanted to dance with each other, but Martha couldn't, because her husband was here.

Whereas the antecedent clause does not contain a VP that denotes the intended referent – i.e., the meaning of *dance with Irv* – this meaning can be transparently inferred from the meaning of the antecedent clause.

The fact that VPE exhibits a series of diagnostic properties it shares with other forms of discourse reference is highly consistent with a referential account. Furthermore, some of these properties result in some of the most serious challenges for IDENTITY accounts of VPE, such as the ability to refer in split-antecedent or exophoric contexts (see Section 2.2.2 for details). It is worth stressing, however, that these arguments have proceeded primarily by analogy: it is widely accepted that pronominal reference does not require an IDENTITY relation between a referring expression and its antecedent, and since VPE resembles pronominal reference in many respects, it stands to reason that VPE is not governed by IDENTITY constraints either. What we do not have is a predictive theory that can distinguish cases of successful reference from unsuccessful ones (particularly in cases of split or inferred antecedents, for example), for either pronominal reference or VPE. While we recognize developing such a predictive theory as an important long-term goal for discourse theorists, this is not the goal we pursue in this chapter. Instead, we will aim to show that the referential approach is the one that is more likely to ultimately prove successful. To establish this, we sought to find a new type of example that is compatible with referential accounts of VPE, but is far out of reach of IDENTITY theories as well as the various rescue strategies posited to reconcile data that are problematic for such accounts.

For this purpose, we turn to the domain of indirect speech acts, which we illustrate first

with an example involving pronominal reference. To set up the context, imagine that Susan and Tom are in their family room, night has just fallen after a warm day, and Susan is standing next to an open patio door at the time of the exchange in (132).

(132) Tom: I'm cold.
Susan: Sure, I'll close it in a second. [*it* = patio door]

Here, Susan has interpreted Tom's statement as an indirect speech act, specifically a request to close the patio door. Susan responds in accordance with this interpretation, and indeed considers the patio door to have been made salient and topical enough by Tom's utterance to be able to refer to it with the pronoun *it*, even though it has not been explicitly mentioned. Importantly, note that the reference may be perceived as slightly degraded – for example, compared to the use of *the patio door* instead of *it* – an intuition that is expected in light of the theory of reference, since a modicum of inferential work is required to accommodate a discourse model referent for the door. But it succeeds nonetheless.

Now consider a discourse of the sort shown in (133).

(133) Spectator: Can I please see that card trick one more time?
Magician: I'm sorry, I can't.

On perhaps the most natural interpretation given the context, the magician's utterance would be intended to mean *I'm sorry, I can't show you the card trick one more time*, and not the one that IDENTITY would predict, i.e., *I'm sorry, I can't see the card trick one more time*. As we will argue below, the interpretation involving a card-trick *performing* event, rather than the card-trick *observing* event introduced by the antecedent VP, lies outside of the explanatory reach of the IDENTITY rescue strategies surveyed in Section 4.1. On the other hand, such a reading is compatible with a referential theory of VPE, in light of the fact that the spectator's utterance

comes with an indirect speech act conveying a request that the magician perform the trick again. Hence, while not linguistically expressed, a performing event might be made accessible enough for subsequent reference, albeit perhaps with some level of degradedness, on analogy with the status of the patio door in example (132).

This sets the stage for the experiment described in the following section, whose purpose was two-fold: to test whether dialogues like (133) indeed make interpretations beyond the antecedent VP available, and to test whether such interpretations are relatively acceptable despite the extreme violation of IDENTITY that they entail. Towards this end, we couple canonical acceptability judgments with a novel experimental task, in which participants paraphrase the ellipsis site.

4.3 Methods

4.3.1 Participants and materials

We recruited 20 self-reported native English speakers from `Amazon.com`'s Mechanical Turk, who were compensated according to the federal minimum wage of $7.25/hour. They were presented with ten two-turn dialogues like the one in (133), repeated in (134) to illustrate the experimental manipulation. The first utterance contained an active-voice VP ($V = \{$*see, get, hear, know, borrow*$\}$), and the second utterance employed auxiliary-focus VPE.

(134) **Spectator:** Can I please see that card trick one more time?

a. **Magician:** I'm sorry, you can't. [NO-CHANGE variant]

b. **Magician:** I'm sorry, I can't. [CHANGE variant]

The first utterance was designed to do two things: to provide a VP that could serve as the antecedent for subsequent ellipsis, and to communicate an indirect speech act by implicitly

submitting a request—in this case, that a card trick be performed. As for the response, we manipulated in a within-item and within-participant design whether the pronominal subject of the ellipsis clause referred to the same person as the antecedent-clause subject (the NO-CHANGE condition) or to a different person (the CHANGE condition). Consequently, the antecedent VP was identical across conditions, but the antecedent-provided referent yielded a plausible interpretation only in the NO-CHANGE condition (*You can't see that card trick one more time* [you = the spectator]), and not in the CHANGE condition (*I can't see that card trick one more time* [I = the magician]).

The ten experimental items were interspersed with 20 non-elliptical and 20 elliptical filler dialogues, and the latter were used to establish upper- and lower-bound baselines for the acceptability and interpretation of run-of-the-mill VPE. Lower-bound elliptical fillers involved violations of various sorts, and lower-bound *non*-elliptical fillers involved extraction violations. Examples of filler items are given in (135) below.

(135) a. **Upper-bound elliptical filler:**

Can you see the remote control anywhere? —I can't, sorry.

b. **Lower-bound elliptical filler:**

Of course I believe in God.

—Even though a proof that God exists doesn't?

c. **Upper-bound non-elliptical filler:**

I can't hear what he's saying. —I don't care.

d. **Lower-bound non-elliptical filler:**

I'm telling you, I didn't get your email.

—That's impossible, whose did you get email?

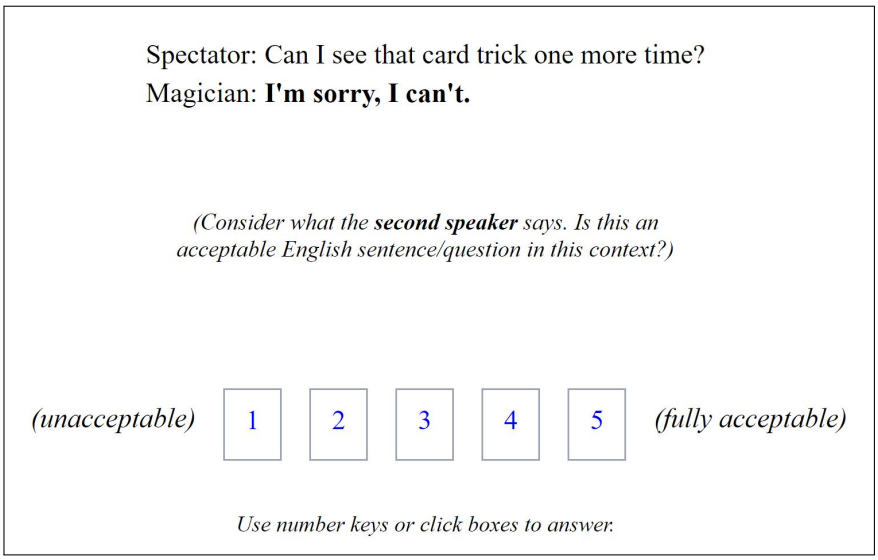
4.3.2 Procedure

Dialogues were presented twice in separate blocks, each time in random order. In the first block, participants were asked to rate the second utterance in terms of its acceptability in context (cf. Fig. 4.1A). In the second block, they were asked to paraphrase the ellipsis site in a free production task (cf. Fig. 4.1B). The purpose of this task was to investigate what interpretations the ellipses received and to what extent those interpretations deviated from the meaning of the antecedent VP. To that end, the paraphrases were analyzed by extracting the main verb and comparing it to the head of the preceding VP to determine whether comprehenders' interpretation of the ellipsis site matched the preceding VP. Reducing this comparison to the verb itself greatly improved the tractability and transparency of the analysis of the free-production results from the paraphrase task.

4.3.3 Analysis and predictions

If comprehenders interpret the dialogues in line with the grammatical constraints of IDENTITY theories, both NO-CHANGE and CHANGE items should be interpreted according to the antecedent-provided meaning. Consequently, we would expect paraphrases of the ellipsis site to use the antecedent verb to express that meaning to similar degrees in both conditions, and we would expect comparable levels of acceptability (modulo a possible reduction in acceptability of the CHANGE variant due to a reduction in plausibility) since items in both conditions observe IDENTITY. If, on the other hand, comprehenders adopt a different interpretation for CHANGE items (e.g. *I can't show you the card trick again* in (134b)), that should be reflected in a lower degree of overlap between antecedent and paraphrase verbs than in the NO-CHANGE condition. That sort of interpretation would entail an extreme mismatch between the antecedent VP and the ellipsis site, which should thus result in low acceptability, perhaps on a par with that of lower-bound elliptical fillers.

A



B

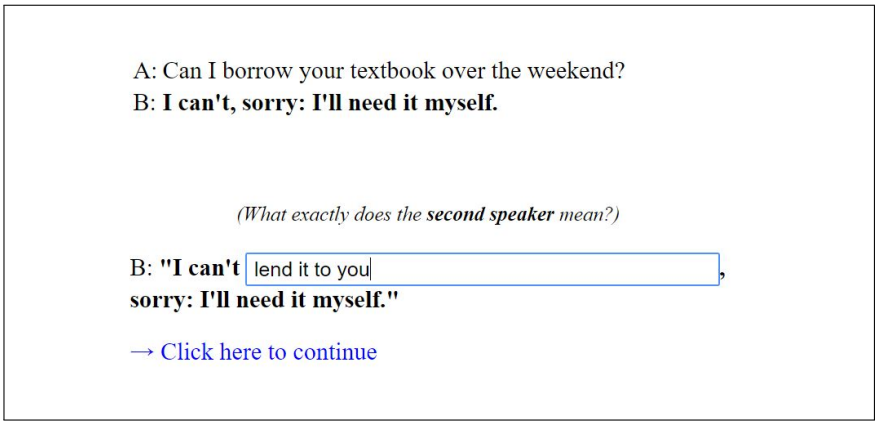


Figure 4.1: Screenshots illustrating the **(A)** acceptability judgment task and the **(B)** paraphrase task. Both examples show CHANGE variants.

The referential theory of VPE makes different predictions. On that view, we expect the ellipsis site to be able to pick up inferred referents if the discourse context is sufficiently conducive. Thus, comprehenders should be free to adopt interpretations that deviate from the antecedent VP to the extent that our experimental items succeed in providing such a conducive context. Any difference in acceptability between CHANGE and NO-CHANGE items would be expected to be in

proportion to the inferential work associated with accommodating the necessary referent. We further analyzed the variability associated with participants' verb choice in the paraphrase task because we suspected that, if comprehenders do infer VPE meanings beyond the one provided by the antecedent VP, those inferred meanings may be less precisely "carved out" than referents that were introduced by purely linguistic means. This prediction reflects the privileged status the referential theory assigns to antecedent-provided referents.

We tested these hypotheses by performing hierarchical regression analyses with Gaussian (for acceptability) and logistic response distributions (for the proportion of antecedent-provided interpretations), comparing CHANGE and NO-CHANGE item variants. Each analysis modeled both the population-level effects corresponding to the hypothesis in question, as well as maximal by-item and by-subject group-level intercepts and slopes (Barr et al., 2013).

4.4 Results

The results are summarized in Figure 4.2. There was a strong tendency for CHANGE items to deviate in their interpretation from the preceding VP: only 5% of paraphrases employed the antecedent verb, compared to 74% for their NO-CHANGE counterparts ($\beta = 4.47$, $p < .001$; see Fig. 4.2B). (The rate of antecedent verb re-use in the NO-CHANGE condition was numerically, but not significantly, lower than that of upper-bound elliptical fillers, whose mean is shown by the dashed line.) Despite this deviation, CHANGE items were much more acceptable than lower-bound elliptical fillers ($\beta = 1.39$, $p < 0.001$), and only slightly less acceptable than their NO-CHANGE counterparts ($\beta = -0.22$, $p = 0.011$; cf. Fig. 4.2A).

Finally, we analyzed the degree of uncertainty in participants' verb choice during the paraphrase task, operationalized as entropy in the by-item distribution of verbs $v \in V$:

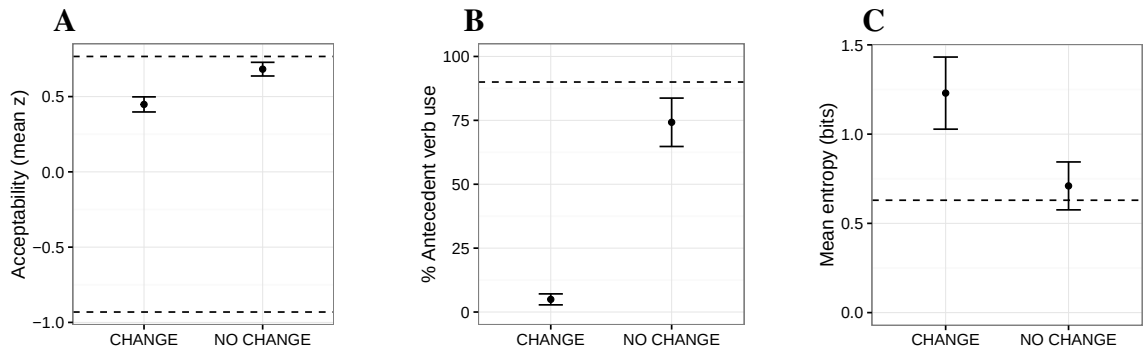


Figure 4.2: Results from the acceptability judgment task (**A**) and paraphrase task (**B** and **C**). **B** shows the proportion of antecedent verb use; **C** shows variation (entropy) in verb choice. Dashed lines show the mean scores for elliptical fillers, exemplified in (135a) and (135b).

(136) $$\mathcal{H}(V) = -\sum_{v \in V} P(v) \log P(v)$$

CHANGE items were associated with significantly more uncertainty in paraphrase verb choice ($t = 2.14, p = 0.048$), although that effect was relatively small (cf. Fig. 4.2C).

4.5 Discussion

Let us first consider the results from the perspective of IDENTITY theories. The results from the paraphrase task show that CHANGE items overwhelmingly received interpretations that deviated dramatically from the antecedent-provided meaning. Since they were nonetheless judged relatively acceptable, our results are problematic for IDENTITY theories in light of the extreme degree of mismatch between antecedent and elided VPs. For example, consider the results for the CHANGE variant of the card trick item discussed earlier. To derive the participants' interpretations under IDENTITY, the elided material must involve either *show* or *do*, both of which featured in roughly 50% of paraphrases:

(137) **Spectator:** Can I please see that card trick one more time?

Magician: I'm sorry, I can't
- ~~show you the card trick again~~. (≈ 50%)
- ~~do the card trick again~~. (≈ 50%)

In both cases, the elided material deviates substantially from the antecedent VP, and together they exhibit two types of mismatch: both involve several lexical mismatches, including the verb itself, and the *show* variant additionally involves a mismatch in transitivity, with ditransitive argument structure in the ellipsis clause where the antecedent clause is transitive. To our knowledge this is the first report of VPE with this degree of mismatch that nonetheless achieves high levels of acceptability.

At this point we may ask whether these cases of non-identity can be accommodated under an IDENTITY analysis with the help of the rescue strategies surveyed in Section 4.1. Let us consider the representational approach first. Recall that this approach would involve reanalyzing all mismatching elements as covertly identical, which in this example will minimally require the following identities: *see* = *show*; transitive = ditransitive. Provided that such a reanalysis is possible at all (it is not clear to us how), this strategy will run into fatal problems once we consider a slightly broader range of data. For example, consider the following item along with the elided material that would be required to derive participants' interpretations under IDENTITY:

(138) **A:** Before Trump got elected, people demanded to see his tax returns, but he refused.
B: And now that he's president, I don't think
he ever will
- ~~release his tax returns~~. (≈ 63%)
- ~~show his tax returns~~. (≈ 27%)
- ~~provide his tax returns~~. (≈ 9%)

If we just focus on the verb, the representational approach to these cases of non-identity would now require not just *see* = *show* = *do*, but also *see* = *provide* and *see* = *release*, which by transitivity would result in unattested identities, such as *do* = *provide*. As a result, such an analysis would grossly overgenerate and, for example, falsely predict that *release the card trick*

should be a possible interpretation of the CHANGE variant of the card trick item.

Inferred-antecedent strategies, such as Thoms' (2015) antecedent accommodation algorithm summarized in (128), face similar challenges when applied to our results. Under such an analysis, the elided material *may* differ from the overt antecedent if it is identical to some member of the set of accommodatable antecedents. Thus, if we allow the necessary lexical substitutions (and/or insertions and deletions) to take place, we may derive the attested interpretations under IDENTITY and explain the high degree of acceptability CHANGE items achieved. Note, however, that permitting the necessary lexical substitutions, such as substituting *show* for *see*, would violate the semantic constraint in (128d) that requires inferred antecedents to be semantically identical to the overt one. Further, other necessary operations, e.g. accommodating a ditransitive antecedent VP from the transitive overt one to derive the *show* interpretation of the card trick item, would further violate the complexity constraint in (128c). These constraints were specifically put in place to prevent overgeneration, and it is easy to see that permitting the operations that would be necessary to derive the attested interpretations would additionally generate a wide range of other interpretations that are unattested.

Thus, the cases of VPE we examined in our experiment represent a class of counterexamples to IDENTITY theories of VPE that neither the representational nor the inferred-antecedent rescue strategy can defuse convincingly. By contrast, the finding that VPE is amenable to accommodated referents in conducive contexts is consistent with a referential account. Particularly the finding that the inferences in question are sensitive to considerations of pragmatic inference and world knowledge is unsurprising given that this is a well-known property of discourse accommodation in other referential domains.

We may ask, however, why CHANGE items were reliably degraded compared to their NO-CHANGE counterparts (Fig. 4.2A). The referential theory we sketched in Section 4.2 predicts that whenever a reference requires accommodating inferences, its felicity depends on the inferential burden comprehenders face in resolving it. Recall in this regard the example provided in (132),

in which the intended referent of the pronoun *it* is made available by an indirect speech act requesting that the patio door be closed. While *it* refers successfully, it may be degraded compared to other referring expressions, such as the definite NP *the patio door*, which would require less inferential work to accommodate the discourse conditions it presupposes. Likewise, while *there* is straightforwardly resolved to *France* in (130), repeated in (139a) below, the analogous inference appears somewhat less straightforward in (139b).

(139) a. Jean is a Frenchman, although he hasn't lived there for many years.

b. Tim is a Dutchman, although he hasn't lived there for many years.

This demonstrates that the felicity of referring expressions that require accommodation depends on the difficulty of performing those inferences in the context (see Section 2.2.3 for more examples and discussion of this point). Since the difference between CHANGE and NO-CHANGE items in our experiment was that the former did require such accommodation while the latter did not, it is not surprising that those inferences incurred some cost in acceptability.

4.6 Conclusion

We have introduced a type of VPE example not previously discussed in the literature in which the meaning of the ellipsis clause is contributed not by the antecedent clause directly, but rather an indirect speech act that it conveys. We have presented experimental evidence that such interpretations do indeed arise when they are supported by the discourse context, and that they achieve relatively high levels of acceptability. This finding is inconsistent with IDENTITY theories of VPE as well as proposed rescue strategies for cases that involve mismatch. Furthermore, our examples exhibit the symptoms of discourse accommodation that are familiar from the literature on reference, and hence point us in the direction of a referential theory of VPE.

These results may seem surprising in light of the fact that, as discussed at the outset of

this chapter, VPE appears to be immune to plausibility-based considerations in many other cases. For example, recall (129), repeated below as (140):

(140) The students had to study hard because the teacher didn't ϕ. [study hard]

Despite the implausibility of the antecedent-provided interpretation and the apparent availability of a more plausible alternative reading, that reading is not a possible interpretation of the ellipsis clause. Whereas this result follows trivially from IDENTITY accounts of VPE, a referential theory has to provide an account for why the necessary inference to accommodate the reference is not supported in this context.

The answer, we believe, bears on a salient disanalogy between cases like (140) and the passages used in our experiment such as (134b). In (140), there are many possible interpretations for the VPE that result in a plausible outcome – *the teacher didn't teach them well, the teacher didn't provide a sample exam, the teacher didn't assure them that the test would be relatively straightforward*, etc. As with any anaphoric expression, context must provide a way of distinguishing a particular interpretation among other possible ones. This is why a sentence like *John said that he studied hard*, spoken discourse initially and outside of a special context (e.g., one in which another male is situationally present), can only be interpreted with *he* referring to John: the addressee is not free to resolve the pronoun to some other male referent, even if plausibility factors point away from John as the referent (e.g., John is notorious for never studying). Similarly, the context in (140) only distinguishes one possible referent from other possible ones—i.e., the one that was linguistically evoked—and hence VPE only succeeds when this is the intended referent.

The situation is different for (134b), however. In this case, the referent required for the non-identity reading is inferred, but it is not inferred in service of satisfying the referential requirements of VPE. It is instead inferred as a by-product of recognizing the spectator's question as an indirect request to show (or do, or perform) the trick. This has the effect of introducing this

potential interpretation into the discourse model as well, with the result that there are now two possible interpretations that are distinguished from all of the others: one introduced by linguistic evocation (*see the trick*), and one resulting from the inference of an indirect speech act (*show the trick*). As our results demonstrate, both are indeed available for subsequent reference with VPE, albeit with the second, inferred interpretation being slightly less accessible.

This observation may explain another type of case that may at first seem puzzling for the referential account, illustrated in (141):

(141) After the test I wasn't sure if I had passed or not.

As it turns out,
- I did. [pass][*fail]
- I didn't. [pass][*fail]
- it's less likely than I thought. [pass][*fail]

Even though the context makes both passing and failing a salient possibility, only the meaning of the antecedent VP—in this case *passing* the exam—is a possible referent for subsequent VPE. While this observation may be interpreted as *prima facie* evidence for IDENTITY accounts of VPE, this cannot be the right story, as the same constraint exists for propositional anaphora: *it* in *it's less likely than I thought* is also limited to the meaning introduced by the antecedent verb. As in the cases surveyed above, in such a context the addressee is faced with the problem of individuating an interpretation from all possible ones, and here one interpretation has been linguistically mentioned, i.e. *passed*. As we would expect, therefore, the judgments reverse if we replace *failed* with *passed* in (141).

To conclude, we have argued that the data examined in this chapter provide substantial evidence for a referential theory of VPE as opposed to IDENTITY accounts. At the same time, they also highlight the importance of further investigating the contextual conditions under which referent can be accessed via VPE. We believe this to be an important goal for future research focused on building increasingly predictive theories of discourse reference in general and of VPE

in particular.

4.7 Note regarding previous publication

The material presented in this chapter was co-authored by Andy Kehler and was published in:

> Poppels, T., & Kehler, A. (2018). Overcoming the identity crisis: Novel evidence for a referential theory of verb phrase ellipsis. In *Proceedings of the Annual Meeting of the Chicago Linguistics Society* (Vol. 53, pp. 403–4017).

Chapter 5

Argument-structure mismatches under sluicing

5.1 Introduction

Recall from Chapter 2 that theories of ellipsis roughly divide into two camps: IDENTITY theories, which posit that ellipsis is grammatical only if the elided material is—in some relevant sense—identical to its antecedent; and referential theories, according to which ellipsis is governed by the same underlying mechanism that enables other forms of discourse reference. This chapter focuses on IDENTITY theories of sluicing by comparing different variants of this approach against a novel set of data. In particular, I will report two experiments on the acceptability of sluicing under mismatches due to tough movement (Experiment 1) and the active/passive voice alternation (Experiment 2). These experiments were designed to speak to different definitions of IDENTITY that have been proposed over time and the results provide new adequacy criteria for theory development in this area. At the end of the chapter, I will briefly discuss the implications of these findings for referential theories of ellipsis.

5.1.1 Constructional mismatches under ellipsis

Languages provide a number of different constructions speakers can choose from to express more or less the same thought, including cleft constructions, argument-structure alternations, and tough movement:

(142) a. No one has solved this problem.
b. This problem hasn't been solved.

(143) a. Someone broke the mirror.
b. The mirror broke.

(144) a. It's easy to elide redundant material.
b. Redundant material is easy to elide.

Such minimal pairs provide informative test cases for IDENTITY theories of ellipsis (Lipták, 2015, see also Chapter 2): since they do not affect the truth-conditional meaning of the utterance but do involve distinct syntactic configurations, constructional mismatches have implications for the level of representation at which the IDENTITY condition is to be defined. According to purely semantic IDENTITY theories (e.g., Merchant, 2001), substituting constructional variants under ellipsis should be inconsequential since doing so does not affect the meaning of the elided material or its antecedent. Syntactic IDENTITY theories (e.g., Chung et al., 1995), on the other hand, would predict any differences in syntactic structure to render ellipsis ungrammatical, and thus constructional mismatches should be categorically unacceptable according to such theories. Therefore, if constructional mismatches were found to render ellipsis unacceptable, that would provide strong evidence that the IDENTITY condition is at least partially syntactic in nature (Chung et al., 2011; Van Craenenbroeck & Merchant, 2013; Lipták, 2015).

The empirical picture, however, turns out to be more complicated than this: at least with respect to VP-ellipsis, there are both acceptable and unacceptable cases of constructional

mismatch. Consider first the following examples, each of which involves a constructional mismatch that appears to render the use of ellipsis infelicitous:[1]

(145) a. The problem was looked into by John, and Bob did #(look into the problem), too. (Kehler, 2000, ex. 34)

b. Even if you want me to shut up, you can't #(shut me up).

c. It's easy to identify venomous snakes, and poisonous plants are #(easy to identify) as well. (Kertz, 2013, p. 407)

By contrast, there are naturally attested mismatch cases as well, and some of them appear to be perfectly felicitous:

(146) a. This problem was to have been looked into, but obviously nobody did (look into the problem). (Kehler, 2000, uttered by Vincent Della Pietra in conversation)

b. And I know that as much as some of you might want me to (shut up), it's 2018 and I'm a woman so you cannot shut me up.[2]

c. Venomous snakes are easy to identify, and most experienced hikers can (identify them). (Kertz, 2013, p. 407)

This state of affairs has spurred a large experimental literature on VP-ellipsis with the goal of understanding the factors behind the variable effect of constructional mismatches (Arregui et al., 2006; Kertz, 2010; Kim et al., 2011; Kertz, 2013; Kim & Runner, 2018; Poppels & Kehler, 2019).

[1]Note that matched variants of these examples are acceptable:

(i) John looked into the problem, and Bob did (look into the problem), too.

(ii) Even if you want to shut me up, you can't (shut me up).

(iii) Venomous snakes are easy to identify, and poisonous plants are (easy to identify) as well.

[2]Michelle Wolf during the 2018 White House Correspondents Dinner, available at the time of writing at https://youtu.be/L8IYPnnsYJw?t=2m26s.

With respect to sluicing, constructional mismatches have received much less attention from experimentalists (but see SanPietro et al., 2012), although their theoretical relevance has long been recognized (Levin, 1982; Merchant, 2008; Tanaka, 2011b; Merchant, 2013b; Lipták, 2015; Rudin, 2019). For example, Merchant (2005), Chung (2006, 2013), Merchant (2008, 2013b) and Rudin (2019) all point to infelicitous argument-structure mismatches, exemplified in (147), as evidence that the IDENTITY condition on sluicing must be (at least partially) defined over syntactic representations.

(147) a. They embroidered something on their jackets, but I don't know with what #(they embroidered their jackets). (Merchant, 2005, ex. 79b)

b. The window suddenly closed, but I don't know who #(closed it).
(Chung et al., 2011, ex. 24)

c. I saw someone's dancing, but I can't remember whom #(I saw dancing).
(Tanaka, 2011a, ex. 115)

d. Someone abducted the candidate, but we don't know by who #(the candidate was abducted). (Chung et al., 2011, ex. 25b)

In the absence of experimental studies involving variations of such examples, however, we do not know whether they reflect categorical grammatical violations (as is mostly assumed in the literature), or whether sluicing is similar to VP-ellipsis in that constructional mismatches exhibit gradience.

The goal of this chapter is to begin to fill this gap in the literature by reporting two experimental studies on constructional mismatches under sluicing. The first experiment involves mismatches based on tough movement. The second experiment addresses the issue of voice-mismatched sluicing. Both experiments reveal novel patterns that have implications for syntactic IDENTITY theories of sluicing and thus expand the empirical base of the literature by contributing novel adequacy criteria for theories of ellipsis.

5.1.2 Implications for IDENTITY theories of sluicing

To appreciate the theoretical significance of constructional mismatches to theories of sluicing, it is useful to step back and briefly review past approaches that focused exclusively on either syntactic or semantic representations in their definition of IDENTITY. Chung et al. (1995) proposed that sluicing is enabled by a process that copies the syntactic representation of the antecedent clause into the ellipsis site, thereby giving rise to a syntactic IDENTITY condition. As outlined in Chapter 2, however, purely syntactic IDENTITY conditions are challenged by a class of acceptable lexical mismatches termed "vehicle change" by Fiengo and May (1994), which are exemplified by the elided pronoun *her* and its antecedent NP *a peaceful protester* in (148).[3]

(148) The cops brutalized a peaceful protester, even though she didn't think they would (brutalize her).

While the elided pronoun and its correlate in the antecedent are co-referential and therefore semantically identical, they are syntactically distinct and thus should render sluicing infelicitous. This problem (among other considerations) later inspired Merchant (2001) to propose a purely semantic IDENTITY condition instead, which is known as "e-GIVENness" and remains one of the most influential theories of sluicing to date (see Section 2.1.1 for details). Restricting the IDENTITY relation to semantic representations alone also has its problems, however, and this is where constructional mismatches come in. As outlined above, purely semantic IDENTITY theories predict such mismatches to be inconsequential, but instead they appear to render sluicing ungrammatical. Consider the following examples of sluicing:

(149) a. Someone tear-gassed the protesters, but the reporters couldn't see who (tear-gassed the protesters). [active-active match]

b. The protesters were tear-gassed, but the reporters couldn't see who #(tear-gassed

[3] See Section 2.1.1 for more details on "vehicle change" mismatches.

the protesters). [passive-active mismatch]

c. The protesters were tear-gassed, but the reporters couldn't see by who (the protesters were tear-gassed). [passive-passive match]

d. The protesters were tear-gassed, but the reporters couldn't see who #(tear-gassed the protesters). [passive-active mismatch]

While (149a) is perfectly felicitous, passivizing the antecedent clause, as in (149b), renders sluicing infelicitous, and an analogous contrast exists between the passive-passive voice-matched example in (149c) and its voice-*mis*matched counterpart in (149d). Since passivization does not affect the truth conditions of the antecedent, purely semantic IDENTITY theories, such as e-GIVENness, cannot explain these contrasts.

To remedy this shortcoming, subsequent proposals have adopted "hybrid IDENTITY" conditions that reference both semantic and syntactic representations (e.g., Chung, 2006, 2013; Merchant, 2013b). For example, Chung (2006, 2013) combined e-GIVENness with a lexico-syntactic condition known as the "No New Words constraint" (sometimes also referred to as "Chung's generalization"), which prevents the ellipsis site from containing any lexical material not provided by the antecedent clause. This condition successfully rules out any constructional mismatches that involve the ellipsis of lexical material not provided by the antecedent, even if they do not violate e-GIVENness. For example, (149b) can be ruled out under the assumption that passive and active variants of verbs are distinct lexical items (a common assumption at least since Hale and Keyser, 1993), as illustrated in (150) (I use underlining to indicate violations of the No New Words constraint):

(150) The protesters were $\underline{\text{tear-gassed}_{\text{PASSIVE}}}$ but they don't know who #($\underline{\text{tear-gassed}_{\text{ACTIVE}}}$ them).

Rudin (2019) follows the same logic in ruling out argument-structure mismatches like

(150): his IDENTITY condition also prevents mismatches between lexical items that project different argument structures (small *v* in his case), and it further imposes a structure-matching constraint that additionally penalizes any differences in word-order that arise from constructional mismatches, which will be discussed in more detail in the context of Experiment 1 below.

Note that both Chung's and Rudin's accounts are explicitly limited to sluicing and applying them "as is" to VP-ellipsis would categorically—and mistakenly—rule out voice-mismatched VP-ellipsis on the same grounds, since the elided predicate (or small *v*) in such examples is lexically or featurally distinct from its correlate in the antecedent:[4]

(151) This problem should have been solved, but obviously nobody did (solve it).

The challenge of providing a unified explanation for the variable effect of voice mismatches on VP-ellipsis and sluicing has prompted (Merchant, 2008, 2013b) to pursue a slightly different angle (see also Tanaka, 2011b). Since Merchant adopts Chung's hybrid IDENTITY account that combines e-GIVENness with the No New Words constraint, voice mismatches are also attributed to a lexical mismatch (between active and passive Voice heads), but the domain of VP-ellipsis is reduced to the VP node so that VoiceP remains unaffected by the IDENTITY requirement and is consequently allowed to vary freely. Sluicing, on the other hand, does not allow such freedom because it involves the ellipsis of an entire clause, including its VoiceP. The difference between VP-ellipsis and sluicing thus derives from differences in the size of the elided constituent and, correspondingly, the domain of IDENTITY, as shown in (152), where we use ~~strike-out~~ font to indicate the domain of IDENTITY according to Merchant (2013b).

(152) This problem should ~~been solved~~, but I don't know...

[4]Rudin (2019) briefly comments on this issue and concedes that extending his analysis to VP-ellipsis would require walking back his "eventive core" generalization precisely because allowing voice-mismatched VP-ellipsis would require re-defining the domain of IDENTITY as strictly smaller than *vP* so as to allow the Voice head to vary freely.

a. *...who [~~$_{TP}$ [**VoiceP** solved the problem]]~~].

b. ...if anyone ever will [**VoiceP** [~~$_{VP}$ solve the problem]]~~].

While this strategy succeeds in providing a unified account of voice mismatches across VP-ellipsis and sluicing, it fails to capture the gradience associated with voice-mismatched VP-ellipsis: since it categorically classifies them as acceptable for the reasons outlined above, it has to attribute the fact that some cases, such as (145a), repeated in (153), are unacceptable to independent factors external to the theory of ellipsis.

(153) The problem was looked into by John, and Bob did #(look into the problem), too. = (145a)

With respect to sluicing, all of the accounts surveyed above predict a categorical distribution: since argument-structure mismatches will invariably violate the No New Words constraint (as well as Rudin's structure-matching condition), they should render sluicing ungrammatical across the board. Indeed, it is widely assumed in the literature that this is the case, and it is the central goal of this chapter to test this assumption experimentally. Our approach to this issue is loosely inspired by the analogous literature on VP-ellipsis, which has found variable mismatch effects for both voice mismatches and mismatches due to tough movement (e.g., Kertz, 2013): Experiment 1 considers sluicing under mismatches due to tough movement and Experiment 2 examines voice mismatches.

5.2 Experiment 1: tough mismatches

The goal of Experiment 1 was to address the question whether sluicing is sensitive to mismatches due to tough movement.

5.2.1 Methods

Stimuli

We constructed 24 items, each with 12 variants according to a 2x2x3 within-item and within-participant design, as shown in (154). We independently manipulated the presence or absence of ELLIPSIS, whether or not there was a MISMATCH, and which WH-WORD was employed (*how*, *when*, *where*). Controlled comparisons of elided and unelided variants face the possibility that unelided utterances may be subject to a "repeated-clause penalty" since comprehenders may expect clauses to be sluiced whenever it is felicitous to do so (Gordon, Grosz, & Gilliom, 1993; Kertz, 2013). To preempt this issue, unelided variants were constructed by reducing redundant material as much as possible, for example by pronominalizing repeated NPs (Kim & Runner, 2018; Poppels & Kehler, 2019).

(154) a. It's easy to replace brake fluid if you know {how|when|where}.
[+ellipsis, -mismatch]

b. Brake fluid is easy to replace if you know {how|when|where}.
[+ellipsis, +mismatch]

c. It's easy to replace brake fluid if you know {how|when|where} to replace it.
[-ellipsis, -mismatch]

d. Brake fluid is easy to replace if you know {how|when|where} to replace it.
[-ellipsis, +mismatch]

Notice that the within-item manipulation of WH-WORD creates a potential issue with respect to plausibility, since all three questions may not be equally plausible in the same context. For example, while a *how* question is highly relevant in (154), it is less clear how the ease of changing brake fluid depends on knowing *when* or *where* to do it. However, due to the orthogonal manipulation of these factors, any such difference would affect both matched and mismatched

Table 5.1: Range of sluice-embedding clauses across experimental items.

sluice-embedding clause	# items
if you know	8
unless you know	5
even if you know	2
and you should figure out	1
as long as you know	1
if someone shows you	1
if you don't know	1
it's not always clear	1
once I figured out	1
unless you know exactly	1
until you figure out	1
without knowing	1

variants and thus will not interfere with our main research question, which is about the effect of mismatch. Indeed, as we will see, the within-item design will enable us to conduct tightly controlled comparisons that will turn out to be highly informative.

In addition to the 2x2x3 within-item manipulation, we varied the matrix clause of the embedded question across experimental items as shown in Table 5.1. The experiment further included 48 filler items (2:1 ratio), which were designed to establish upper- and lower-bound baselines and distract from the purpose of the experiment. To that end, half of the items were non-elliptical and both elliptical and non-elliptical fillers included acceptable and unacceptable sentences, exemplified in (155).

(155) a. Betsy did after Peter went to the store. [+ellipsis, -acceptable]

b. The thief was arrested and his brother was as well. [+ellipsis, +acceptable]

c. Who did the press secretary ask a question before we interviewed? [-ellipsis, -acceptable]

d. Sometimes Susan has a hard time keeping up in class. [-ellipsis, +acceptable]

Participants and procedure

We recruited 43 participants via Amazon.com's crowd-souring platform Mechanical Turk. Each participant was presented with one variant of each of the 24 experimental items, interspersed with fillers. On each trial, participants judged whether "the sentence was an acceptable English sentence and whether they could imagine themselves or other native speakers saying it," on a scale from 1 ("unacceptable") to 5 ("acceptable"). Two participants were excluded from the analysis because they identified as non-native speakers of English at the end of the experiment. We further excluded all trials with response times below 1000 ms (a total of 592 observations) under the assumption that it is not possible to carefully read and judge the experimental materials in less than a second, leaving us with a total of 2394 observations from 41 participants to analyze.

5.2.2 Results

Population-level averages are shown in Figure 5.1. Inspecting these averages, two patterns emerge. First, there does not appear to be a consistent mismatch penalty: the horizontal lines connecting match and mismatch conditions are not consistently downward sloping. Secondly, while *how* variants appear to be at ceiling both with and without ellipsis, their *when* and *where* counterparts appear to be degraded, especially under ellipsis.

To test whether these patterns are statistically robust, we performed two multilevel cumulative probit regression analyses, both with population-level effects for each condition in the 2x2x3 design (including all interactions) and crossed group-level effects for items and participants including individual intercepts and slopes for all population-level effects. The first model was designed to test our primary hypothesis, i.e. whether sluiced questions were less acceptable in the mismatch condition than in the matched condition. In order to average over WH-WORD, this 3-way factor was sum-coded, whereas MISMATCH and ELLIPSIS were both treatment-coded with MATCH and ELLIPSIS as respective baseline values. This analysis revealed

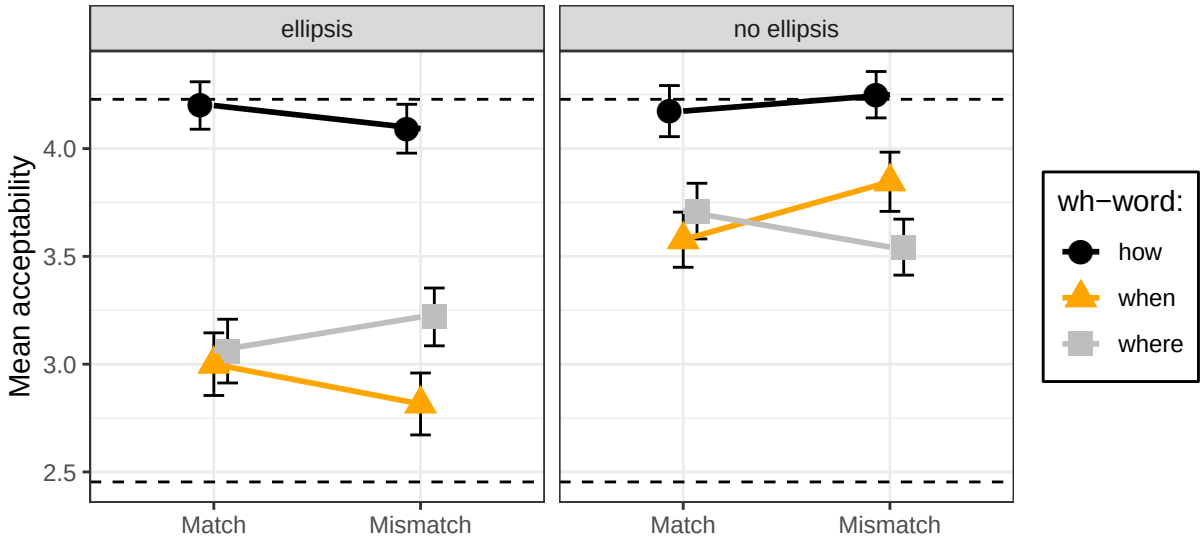


Figure 5.1: Condition averages from Experiment 1.

no evidence for a mismatch penalty for either elided variants ($\Delta = 0.01$, $CI(\Delta) = [-0.24, 0.26]$, $P(\Delta < 0) = 0.47$) or unelided variants ($\Delta = 0.12$, $CI(\Delta) = [-0.16, 0.42]$, $P(\Delta < 0) = 0.2$), and the ELLIPSIS:MISMATCH interaction was also non-significant ($\Delta = 0.11$, $CI(\Delta) = [-0.23, 0.47]$, $P(\Delta < 0) = 0.27$). The second model was designed to test whether *when* and *where* items were degraded compared to their *how* counterparts, and if so, whether that effect was exacerbated under ellipsis. For that purpose, MISMATCH was sum-coded while ELLIPSIS and WH-WORD were treatment coded (baseline values: NO ELLIPSIS and HOW). This analysis revealed that unelided *when* and *where* questions were indeed significantly less acceptable than the corresponding *how* variants (*when*: $\Delta = -0.73$, $CI(\Delta) = [-1.12, -0.34]$, $P(\Delta < 0) = 1$; *where*: $\Delta = -0.84$, $CI(\Delta) = [-1.33, -0.33]$, $P(\Delta < 0) = 1$), and that this effect was significantly exacerbated under ellipsis (*when*: $\Delta = -0.82$, $CI(\Delta) = [-1.40, -0.22]$, $P(\Delta < 0) = 0.99$; *where*: $\Delta = -0.53$, $CI(\Delta) = [-1.02, -0.03]$, $P(\Delta < 0) = 0.98$).

The within-item manipulation of WH-WORD allowed us to further investigate the degradation associated with *when* and *where* questions in a series of posthoc analyses. In particular, since

we modeled item-specific effects with "shrinkage" towards the population-level effects,[5] we were able to explore whether the *when/where* penalties differed across items without manually adjusting for multiple comparisons (Gelman, Hill, & Yajima, 2012). Figure 5.2 shows the item-specific effects of *when* and *where* for both elided and unelided variants. While most items show robust evidence in line with the population-level effects (i.e., most coefficients are negative and elided variants exhibit a greater penalty than unelided variants), there is also some variability across items, especially regarding *where* questions.

To get a sense of what may be driving the penalties, we conducted a qualitative posthoc analysis by inspecting the three items that exhibited the largest and smallest *when/where* degradation. Consider first the following *when* questions, which showed the least evidence for a penalty relative to their respective *how* counterparts.[6]

(156) a. Software updates are important to install but it's not always clear when. (Item 5)

b. A full lunar eclipse is hard to take a picture of unless you know exactly when. (Item 6)

c. Banks are virtually impossible to rob unless you know when. (Item 21)

In all of these examples, the context is such that the *when* question is a plausible continuation. That is particularly clear compared to the three *when* items that exhibited the greatest penalty relative to their *how* counterparts:

(157) a. Some soccer teams are easy to defend against if you know when. (Item 4)

b. Fleas can be hard to get rid of even if you know when. (Item 8)

[5]The term *shrinkage* refers to a property of hierarchical statistical models, whereby population-level effects provide the prior distribution for group-level effects, such as by-item and by-participant random effects. As a result, group-level effect coefficients are biased ("shrunk") towards the corresponding population-level effects and thus more conservative compared to non-hierarchical models, which permits multiple hypothesis tests without inflated significance thresholds.

[6]While we list mismatch variants here, note that the items were selected based on the model coefficients shown in Figure 5.2, which represent the across-the-board difference between *when/where* variants and their *how* counterparts, i.e. averaging over matched and mismatched variants.

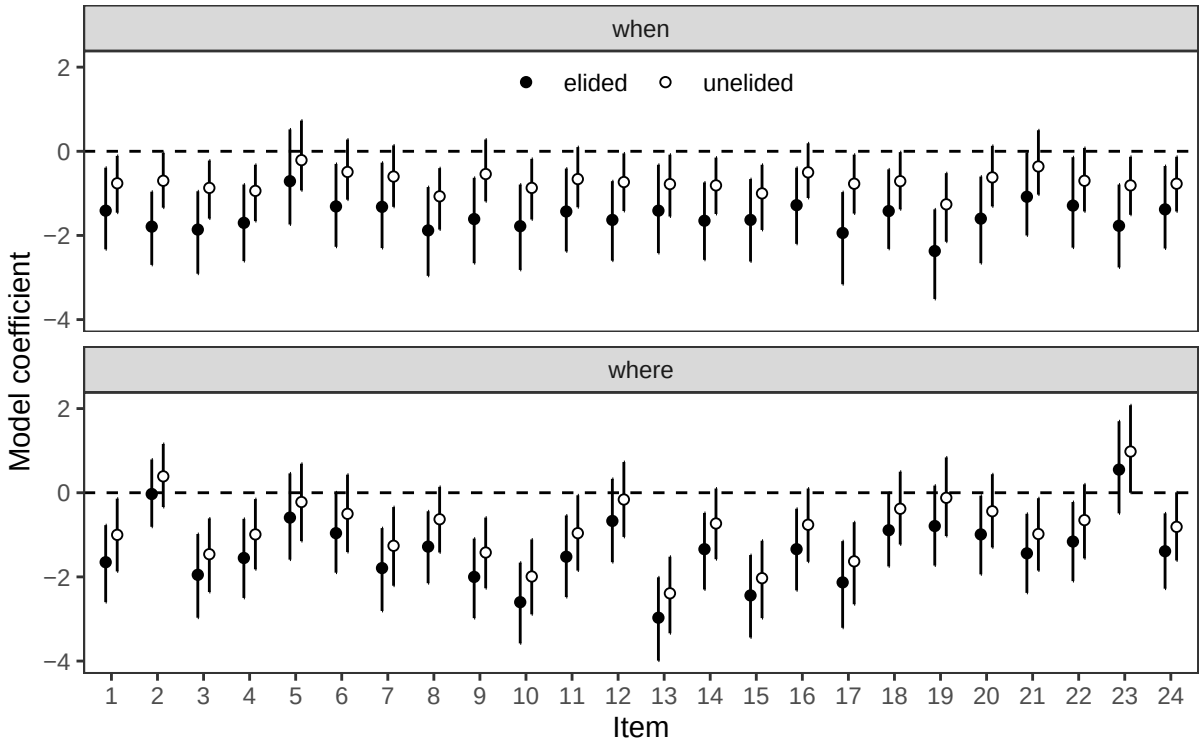


Figure 5.2: Item-specific model coefficients corresponding to the penalties associated with *when* (top) and *where* (bottom) questions relative to the *how* variant of the same item. Black dots indicate elided variants; white dots represent unelided variants. Errorbars show 95% Credible Intervals (i.e., 95% of the posterior samples fall within that interval).

c. Pants that fit perfectly can be impossible to find unless you know when. (Item 19)

Clearly, the embedded questions are less plausible continuations in (157) than in (156), suggesting that the effect in question may be driven by question plausibility. Item 19 is particularly informative in that regard because its *when* variant—shown in (157c)—was significantly degraded whereas its *where* variant—shown in (158a)—was not.

(158) a. Pants that fit perfectly can be impossible to find unless you know where. (Item 19)

b. The truth is that even rare minerals aren't hard to find if you know where. (Item 2)

c. Large cars are almost impossible to park downtown unless you know where.

(Item 23)

These items were not degraded, which once again is consistent with the analysis that the *when/where* degradations are driven by question plausibility. Indeed, Item 23, shown in (158c), exhibits the opposite effect whereby the *where* variant was more acceptable than the *how* variant. This is consistent with the fact that knowing *where* to park large cars is a more plausible bottleneck to parking them downtown than knowing *how* to do so.

Finally, the items exhibiting the largest *where* penalty are shown below, and once again question plausibility appears to play a role:

(159) a. Angry customers are difficult to appease unless you know where. (Item 10)

b. Science can be challenging to explain to children even if you know where. (Item 13)

c. This crime was easy to solve once I figured out where. (Item 15)

While we caution against over-interpreting this posthoc analysis, we take it to provide tentative evidence that the overall degradation of *when* and *where* questions may reflect the effect of question plausibility. Since the penalty in question was exacerbated under ellipsis, we may wonder whether the ellipsis-specific degradation was also due to question plausibility. To assess whether elided and unelided variants were impacted by the same underlying factor, we correlated the respective model coefficients across items. As shown in Figure 5.3, there was indeed a high correlation between the two, especially for *where* items (*when*: $r = 0.85$; *where*: $r = 0.995$).

5.2.3 Discussion

The key findings from Experiment 1 were twofold: we found no evidence that movement negatively affects the acceptability of sluicing; and *when* and *where* sluices were significantly degraded compared to their *how* counterpart, which may reflect a penalty for implausible questions

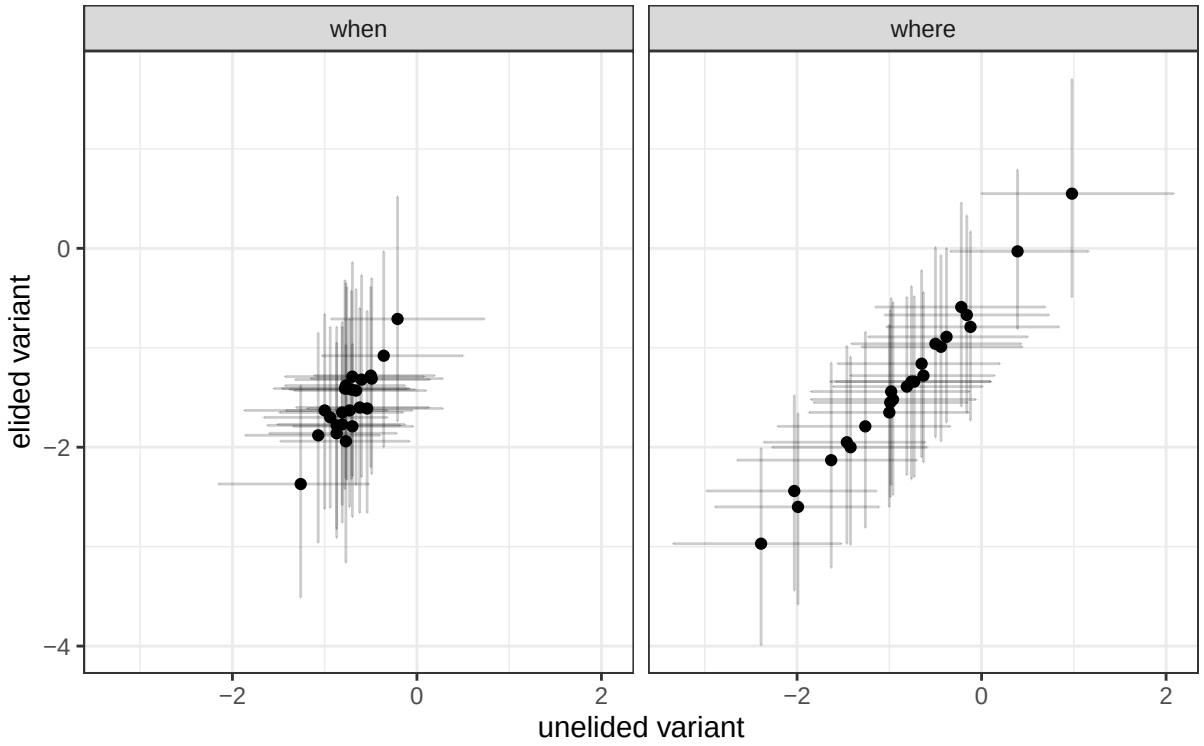


Figure 5.3: Scatterplots showing correlation between item-specific *when/where* penalties associated with unelided (x) and sluiced (y) variants. Errorbars show Credible Intervals around model coefficients.

that is particularly harmful under ellipsis. We will return to the "question plausibility effect" in the General Discussion and focus our present discussion on the implications of the mismatch results for IDENTITY theories of sluicing. According to contemporary theories of syntax, tough movement results in a syntactic trace being left behind after fronting the object (e.g., Messick, 2012), which results in a mismatch between the elided object NP (in this case, *banks*) and the corresponding element in the antecedent clause:

(160) Banks$_i$ are virtually impossible to rob t_i unless you know when ~~to rob banks$_i$~~.

Most contemporary IDENTITY theories of sluicing predict more or less straightforwardly that this mismatch should not affect the acceptability of sluicing. First, it follows trivially from purely semantic IDENTITY theories, such as Merchant's (2001) e-GIVENness condition: since tough

movement does not affect the truth-conditional meaning of the antecedent clause, it also does not affect whether it is in a mutual-entailment relation with the elided material.

Secondly, Chung's (2006) lexico-syntactic "No New Words" constraint is similarly unaffected by tough movement since it is by definition insensitive to word order: while it bans ellipsis sites from containing any lexical material not provided by the antecedent, it does not care where in the antecedent the relevant lexical items are located. The elided NP *banks* is thus licensed by the fronted NP in the antecedent, not by the trace it leaves behind. Therefore, Chung (2006)—as well as other "hybrid" IDENTITY accounts that have adopted this condition (AnderBois, 2010; Merchant, 2013b; AnderBois, 2014)—correctly predict that tough movement should not affect the acceptability of sluicing.

Third, according to the "limited syntactic IDENTITY" account proposed by Chung (2013), the syntactic IDENTITY requirement for sluicing is reduced to two conditions, neither of which applies to the tough movement cases in question: a case-matching condition, which only applies to DP remnants; and an argument-structure condition, which only applies to remnants that serve as internal arguments to an elided predicate. The fact that our experimental materials exclusively feature *how*, *when* and *where* remnants renders them exempt from this "limited syntactic IDENTITY" condition and sluicing should therefore be possible despite tough movement.

Finally, Rudin's (2019) syntactic IDENTITY condition is more restrictive than the other theories mentioned so far. For example, he requires structure matching in addition to lexical identity, which requires the elided object *banks* to be compared to the trace in the antecedent clause, instead of the fronted NP. However, Rudin's definition of lexical identity includes an explicit exception for lexically distinct elements that are syntactically co-indexed. This stipulation is an extension of Fiengo and May's (1994) notion of *vehicle change*, which, following Merchant (2001), is motivated by examples like the following:

(161) I don't know who$_1$ t_1 said what$_2$, or why ~~they$_1$ said it$_2$~~. (Rudin, 2019, ex. 19a)

Just as in the cases involving tough movement, the ellipsis clause contains lexical items that are distinct from, but syntactically co-indexed with, their structure-matched correlates in the antecedent clause: *they* and *it*. With the help of this "vehicle change" provision, then, Rudin (2019) also derives the acceptability of sluicing under tough movement.

In summary, Experiment 1 found that sluicing is insensitive to mismatches due to tough movement, which reifies the need for syntactic IDENTITY theories to either ignore structure-matching violations—as in Chung (2006, 2013) and Merchant (2008, 2013b)—or to carve out a "vehicle-change" exception with respect to syntactic traces, along the lines of Rudin (2019). Zooming back out to the level of comparing sluicing and VP-ellipsis, however, a curious picture is emerging. While both tough movement and voice mismatches lead to similar violations with respect to VP-ellipsis (Kertz, 2013), sluicing reveals a dissociation between the two: it appears to be rendered unacceptable by voice mismatches, but it is unaffected by tough movement. To better understand this dissociation, we now take a detailed look at voice mismatches.

5.3 Experiment 2

5.3.1 Motivation

While Chung (2006, 2013), Merchant (2008, 2013b), and Rudin (2019) all attribute the impossibility of sluicing under voice mismatches to whatever lexical items encode Voice, Merchant's (2013b) additionally focuses on explaining the fact that VP-ellipsis is less sensitive to voice mismatches. With respect to sluicing, however, all of these accounts predict that voice mismatches should be ruled out categorically. The goal of Experiment 2 is to test this prediction.

Because sluicing remnants are minimal compared to the remnants in VP-ellipsis, it is impossible to "force" a voice mismatch whenever the remnant wh-phrase does not include the *by* phrase. As a result, we cannot know *a priori* whether examples like (162) involve a voice mismatch, since both voice-matched and -mismatched parses are compatible with the remnants

of the ellipsis clause.[7]

(162) The problem has finally been solved but I don't know...

a. ...how (it has been solved). [voice-matched]

b. ...how (they solved it). [voice-*mis*matched]

In our experiment, we address this concern in two ways. First, we design the materials in a way that makes the passive-voice interpretation implausible, as shown in (163).

(163) The problem hasn't been solved because no one knows...

a. ...how #(the problem hasn't been solved).

b. ...how ?(to solve it).

While the ellipsis clause can in principle be interpreted in a way that preserves syntactic IDENTITY, we argue that doing so leads to an implausible construal of the passage as a whole due to the causal connective *because*: the reason that the problem in question hasn't been solved is not that no one knows how it hasn't been solved *in the past*, but rather how to solve it *in principle*.

In addition to this plausibility-based precaution, we conducted a separate norming experiment to confirm which interpretations participants did, in fact, adopt, which is described in more detail below.

5.3.2 Method

Stimuli

We created 12 experimental items that followed a 2x2x3 design that independently varied the presence/absence of ELLIPSIS and MISMATCH within items, as shown in (164).

[7] See Chung (2013) for a similar point.

(164) a. The problem has never been solved because no one knows how.

[+ellipsis, +mismatch]

b. Nobody ever solved the problem because no one knows how.

[+ellipsis, -mismatch]

c. The problem has never been solved because no one knows how to solve it.

[-ellipsis, +mismatch]

d. Nobody every solved the problem because no one knows how to solve it.

[-ellipsis, -mismatch]

We additionally varied WH-WORD (*how*, *when*, and *where*), but in contrast with Experiment 1, this manipulation was applied between items in order to have more fine-grained control over the plausibility of the voice-matched readings that needed to be ruled out.

In addition to the 12 experimental items, participants were presented with 12 fillers items designed for a separate experiment that need not concern us here.

Participants and procedure

We recruited a total of 52 participants from Amazon.com's Mechanical Turk platform. 5 participants were excluded for self-identifying as non-native speakers of English, and an additional 198 individual trials were excluded for lasting less than 1,000 ms, leaving us with 930 observations from 47 participants. The procedure was identical to that of Experiment 1: participants judged the acceptability of experimental items, which were interspersed with fillers, on a scale from 1 ("unacceptable") to 5 ("acceptable").

Norming experiment

In order to verify that comprehenders did indeed adopt the active-voice parse that was critical to our MISMATCH manipulation, we recruited a separate set of 34 participants to participate

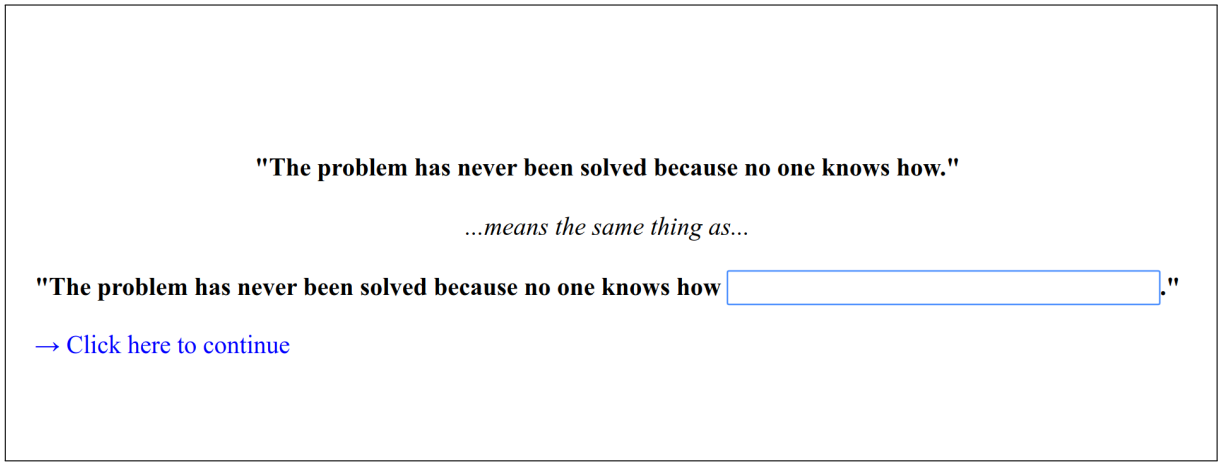


Figure 5.4: Screenshot of a sample trial during the norming experiment. Participants first read the elliptical utterance and then used a free-response text box to indicate their interpretation of the ellipsis site.

in a norming experiment. 13 participants either reported being non-native speakers or submitted clearly bot-like responses[8] and were excluded from the analysis along with any individual trials that took less then 3,000 ms, under the assumption that it is not feasible to process the sentence carefully and paraphrase the ellipsis site in less then 3 seconds. The remaining participants were presented with MATCH or MISMATCH variants of all experimental items and were asked to paraphrase the ellipsis site, as shown in Figure 5.4. We then hand-annotated each response in terms of three categories: active-voice responses; passive-voice responses; and "other," which included cleft completions (e.g., "it was"). 206 of the total 252 responses reflected straightforward active-voice interpretations (81.7%), 38 fell into the "other" category (15.1%), and only 8 responses used passive voice (3.2%). It thus appears that the plausibility manipulation was successful in swaying comprehenders away from adopting passive-voice parses of the ellipsis clause.

At this point, it is worth addressing a potential concern one might have about the norming experiment: does the paraphrase task provide reliable evidence as to participants' parse of the ellipsis clause? Specifically, can we be sure that the distribution of responses does not reflect constraints on *un*elided utterances, which would be the case if participants interpreted the task as a

[8]For example, some responses consisted of language copied from the instructions, such as "Please try to capture the meaning of the second sentence as precisely as possible."

mere "passage completion" task and ignored the elliptical utterance. There are two considerations that we believe alleviate this concern. First, paraphrase tasks of this sort have been used to probe the structure of the ellipsis site before. For example, Frazier and Duff (2019) argue that comprehenders' paraphrases of elliptical utterances are likely to re-use the syntactic material they infer when resolving the ellipsis. Secondly, even if some proportion of participants did adopt a passive-passive parse in the MISMATCH condition, the materials were specifically designed so that they would either be stuck with an IDENTITY-preserving but highly implausible interpretation, as in (165a), or else have to contend with additional lexical mismatches, as exemplified in (165b).[9]

(165) The problem has never been solved because...

a. ...no one knows how (# the problem has never been solved).

b. ...no one knows how (the problem can be solved).

Taken together, these considerations support the assumption that the majority of experimental participants presented with the stimuli in Experiment 2 did adopt the passive-active parse that the MISMATCH condition was designed to test.

5.3.3 Results

Population-level averages are shown in Figure 5.5. Upon visual inspection of these averages, two patterns emerge. First, there does not appear to be a robust mismatch penalty (horizontal lines are not consistently downward-sloping). Secondly, sluiced questions appear to be somewhat degraded compared to unelided variants, especially for *when* and *where* items. To test these two observations statistically, we fit a multi-level model according to the 2x2 design of the experiment, with sum-coding for both ELLIPSIS and MISMATCH so that main effects can be

[9]While Rudin explicitly permits mismatches above the highest elided small *vP* (the "eventive core"), (164b) does violate less forgiving IDENTITY conditions, including those adopted by Chung (2006, 2013) and Merchant (2008, 2013b).

interpreted as "across-the-board" effects) and added WH-PHRASE as a grouping factor alongside items and participants.[10] This was done for three reasons. First, in contrast with Experiment 1, we did not vary WH-PHRASE *within* items because we needed precise control over question plausibility in order to rule out voice-matched interpretations (see discussion above). Secondly, none of our primary research questions or the predictions from theories we were aiming to evaluate differed across WH-WORD. Finally, modeling WH-WORD as a grouping factor allowed us to perform multiple posthoc hypothesis tests without adjusting for multiple comparisons, since group-level effects are hierarchically related to and thus "shrunk" towards the corresponding population-level effects (Gelman et al., 2012).

The results revealed that all population-level effects were non-significant: there was no "across-the-board" mismatch penalty ($\Delta = 0.07$, $CI(\Delta) = [-0.98, 1.18]$, $P(\Delta < 0) = 0.39$); no overall ellipsis penalty ($\Delta = -0.63$, $CI(\Delta) = [-3.27, 2.25]$, $P(\Delta < 0) = 0.78$); and no interaction between the two ($\Delta = -0.09$, $CI(\Delta) = [-3.25, 2.94]$, $P(\Delta < 0) = 0.54$). Due to the hierarchical structure of the model, we were then able to repeat each of these hypothesis tests for each WH-WORD without having to manually adjust for doing multiple comparisons. This analysis revealed that the mismatch penalty was robustly non-significant across *how*, *when*, and *where* questions (*how*: $\Delta = 0.09$, $CI(\Delta) = [-0.31, 0.51]$, $P(\Delta < 0) = 0.31$; *when*: $\Delta = 0.1$, $CI(\Delta) = [-0.28, 0.49]$, $P(\Delta < 0) = 0.29$; *where*: $\Delta = 0$, $CI(\Delta) = [-0.41, 0.38]$, $P(\Delta < 0) = 0.48$. As we had suspected on the basis of Figure 5.4, however, there was a significant ellipsis penalty for *where* and *when* questions, but not for *how* questions (*where*: $\Delta = -1.5$, $CI(\Delta) = [-2.08, -0.87]$, $P(\Delta < 0) = 1$; *when*: $\Delta = -0.49$, $CI(\Delta) = [-1.03, 0.02]$, $P(\Delta < 0) = 0.97$; *how*: $\Delta = -0.19$, $CI(\Delta) = [-0.81, 0.37]$, $P(\Delta < 0) = 0.75$). Finally, there was no evidence for an ELLIPSIS:MISMATCH interaction for *how* questions $\Delta = -0.29$, $CI(\Delta) = [-1.28, 0.67]$, $P(\Delta < 0) = 0.72$, and only weak evidence for *when* and *where* questions, albeit in opposite directions (*when*: $\Delta = 0.67$, $CI(\Delta) = [-0.37, 1.65]$, $P(\Delta > 0) = 0.91$; *where*: $\Delta = -0.68$, $CI(\Delta) = [-1.66, 0.31]$, $P(\Delta < 0) =$

[10] As usual, all group-level intercepts and slopes corresponding to the 2x2 population-level effect structure were added.

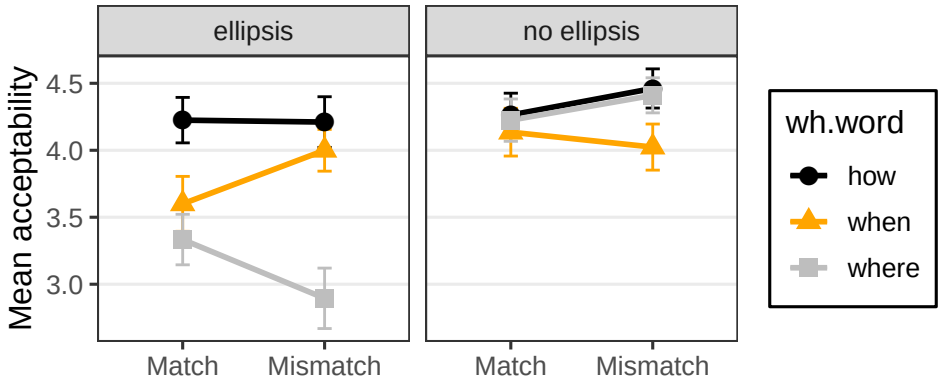


Figure 5.5: Condition averages from Experiment 2.

0.91).

5.3.4 Discussion

With respect to our main research question, the key result from Experiment 2 is that we found no evidence for a mismatch penalty: there was no overall (population-level) effect of mismatch, nor did we find any significant WH-WORD-specific mismatch penalties. This result runs counter to the predictions from syntactic IDENTITY theories (except for Chung (2013) to which we return momentarily): Chung (2006), Merchant (2008, 2013b), and Rudin (2019) all predict that the lexical mismatch between Voice-encoding elements (predicates, small *v*, or Voice head) should render all sluices in the MISMATCH condition ungrammatical. It is worth emphasizing that this mismatch occurs inside the elided TP (indeed, inside its "eventive core"), ensuring that it is subject to the IDENTITY condition in both Merchant's and Rudin's systems, and that the passivization of the antecedent clause additionally incurs a violation of Rudin's structure-matching constraint.

Chung's (2013) "limited syntactic IDENTITY" condition fares better against our results: according to her account, only a small subset of elided syntactic heads are subject to IDENTITY and our items do not include any heads of that sort. Specifically, she only prohibits non-identical

heads that either take the sluicing remnants as an argument or assign Case to a remnant DP, neither of which is the case in our experimental materials. In this way, Chung captures the dissociation between the *how*, *when*, and *where* sluices we examined—which showed no mismatch penalty—and argument-targeting sluices like (166)—which do exhibit a mismatch penalty: since the remnant *who* is an argument of the elided predicate *murdered*.ACTIVE, the argument-structure mismatch with *murdered*.PASSIVE does violate the "limited" IDENTITY condition proposed by Chung (2013).

(166) Joe was murdered but we don't know who #(murdered him).

As we will see in the next section, even Chung's limited syntactic IDENTITY condition turns out to be too restrictive in light of other examples, but it does successfully allow the kind of voice-mismatched sluices that we studied in Experiment 2.

5.4 General Discussion

The main finding across both experiments is the absence of a mismatch penalty: neither tough movement (Experiment 1) nor passivization (Experiment 2) resulted in lower acceptability judgments compared to variants with no mismatch. While all syntactic IDENTITY theories we have considered thus far predict this result with respect to tough movement, only Chung (2013) is consistent with the voice-mismatch results (see Table 5.2): not only does the voice mismatch in our materials violate the lexical-identity requirement of Chung (2006) and subsequent accounts that have adopted it, it also violates Rudin's (2019) structure-matching constraint due to the word-order differences that result from passivization. Finally, the absence of a penalty for voice mismatches refutes the influential line of theories that attribute the variable effect of voice mismatches across different types of ellipsis to the size of the elided constituent (Merchant, 2008; Tanaka, 2011b; Merchant, 2013b): despite the fact that VoiceP is elided in our experimental

Table 5.2: Cross-tabulation of empirical findings and theories of sluicing that require syntactic IDENTITY.

Finding	Chung (2006)	Merchant (2013b)	Chung (2013)	Rudin (2019)
Acceptable tough mismatches (Expt. 1)	✓	✓	✓	✓
Acceptable voice mismatches (Expt. 2)	✗	✗	✓	✗
when/*where* penalty (Expts. 1 & 2)	✗	✗	✗	✗

materials, it can nonetheless deviate from its correlate in the antecedent without incurring an acceptability penalty.

The only account that is consistent with the fact that voice-mismatched sluices were fully acceptable is Chung's "limited IDENTITY" condition. Since it is restricted to syntactic heads that either assign Case to the remnant wh-phrase or take it as an argument—neither of which is the case with our *how*, *when*, and *where* sluices—the IDENTITY requirement is reduced to semantic equivalence (she adopts Merchant's, 2001, e-GIVENness condition).

However, despite being much less restrictive than the other syntactic IDENTITY accounts we considered, Chung's (2013) account is nonetheless too restrictive with respect examples like (167), which is due to Thoms (2015).

(167) I remember someone complaining, but I can't remember who (complained).

Thoms points out that this example appears to be fully acceptable despite the fact that it fails Chung's Case condition: the remnant wh-phrase *who* is assigned Case by the elided finite T head, but since the antecedent clause is non-finite, there is no corresponding head it is identical to. Based on this empirical short-coming (along with considerations of theoretical parsimony), Thoms rejects the notion that syntactic IDENTITY is restricted to a subset of "special heads" in the ellipsis clause, and instead advocates for a "Scope Parallelism" requirement. Interestingly,

while his account, like Chung's, permits the kinds of voice mismatches we found in Experiment 2, it rules out mismatches due to tough movement: the Scope Parallelism requirement has similar consequences as Rudin's (2019) structure-matching condition with respect to word order, but whereas Rudin allows lexically distinct items to count as "identical" if they are syntactically co-indexed, Thoms explicitly prohibits such equivalencies.[11] We are thus in a situation where no account captures the full range of mismatch patterns under consideration: while Thoms (2015) improves on Chung (2013) with respect to cases like (167), he incorrectly rules out the tough movement cases from Experiment 1.

Beyond the absence of a mismatch penalty, both of our experiments also found that *when* and *where* sluices were degraded (an ellipsis-specific effect!), which is an observation that remains mysterious under any syntactic IDENTITY account. This is most obvious in the context of Experiment 1: recall that the stimuli for that experiment were specifically designed to vary the wh-phrase while holding both the antecedent clause and the content of the ellipsis site constant. Consequently, any theory that focuses exclusively on the relation between the elided material and its antecedent will necessarily fail to account for the effect of this manipulation. Furthermore, if the posthoc analysis described above is on the right track, the *when/where* penalty is driven by question plausibility: in contexts that made the relevant question reasonably plausible, the penalties were attenuated, and the most severe penalties were observed in contexts that rendered the to-be-elided question irrelevant. While more research is clearly necessary to verify the role of plausibility, it is worth emphasizing that appealing to pragmatic factors that are external to the theory of ellipsis is not going to be sufficient:[12] While the penalty in question affected both elided and unelided variants in a way that was correlated within items, it was significantly exacerbated

[11] This prohibition results from the "complexity constraint" on the syntactic inference algorithm that Thoms uses to define IDENTITY: lexical mismatches (between semantically equivalent elements) are allowed only if the elided element is at most as complex as its correlate. Since tough movement leaves a trace in the antecedent whereas the ellipsis clause contains a full NP or a pronoun, IDENTITY is violated and ellipsis should be impossible. Note that this conclusion rests on the assumption that tough movement involves A'-movement (see Messick, 2012, for arguments that it does) since Thoms' Scope Parallelism condition is defined so as to be insensitive to A-movement.

[12] See Rudin (2019) for a proposal along those lines.

under sluicing, which suggests that whatever mechanism is responsible for it must be interacting with, rather than operating independently of, the mechanisms that support sluicing. Furthermore, this pattern is also beyond the reach of various "repair" strategies (e.g., Arregui et al., 2006; Frazier, 2013) since there is no grammatical violation to trigger such mechanisms (recall that the *when*/*where* penalty applied to both MATCHED and MISMATCHED variants).

As argued in detail in Chapter 2, variable acceptability profiles are in principle expected under referential theories of ellipsis, since non-elliptical forms of discourse reference exhibit similar gradience. The fact that the variability in the results reported here appear to be driven by question plausibility receives a particularly straightforward explanation under a referential approach: since the use of other forms of discourse reference is heavily influenced by world knowledge and plausibility considerations, the *when*/*where* penalty we found across both experiments is unsurprising if sluicing is governed by the same underlying mechanism. That being said, it is important to emphasize that our analysis of the by-item variability in terms of plausibility is *post-hoc* and allows for tentative conclusions only. Future research could manipulate and quantify question plausibility directly in order to verify that it is indeed causally related to the acceptability of sluicing.

Finally, the results presented in this chapter underscore the value of experimental research in the study of ellipsis. First, while the absence of a mismatch penalty emerged with clarity from our experiments, the ellipsis-specific degradation of *when* and *where* questions could have been misinterpreted as reflecting a mismatch penalty in the absence of experimental control items and careful statistical analysis. Secondly, the ability to compare exploratory findings across items revealed a promising avenue for future research with respect to the *when*/*where* degradation in Experiment 1: since the extent of this penalty was correlated across elided and unelided variants and appears to a first approximation to be associated with the plausibility of the to-be-elided question, this may suggest that theories of sluicing must allow for ellipsis-specific plausibility effects. While more research is necessary to explore this hypothesis, it reflects the benefit of

experimental work on ellipsis.

5.5 Note regarding co-authorship

The material presented in this chapter reflects collaborative work with Andy Kehler and is presently being prepared for publication.

Chapter 6

Sluicing with inferred referents

6.1 Introduction

Before diving into the last chapter of this thesis, it is worth reviewing the contributions of the previous chapters. In Chapter 2 I argued that theories of ellipsis fall into two broad categories: IDENTITY theories that posit that ellipsis is only grammatical if the elided material is identical—in some relevant way—to its linguistic antecedent, and referential theories according to which ellipsis sites contain null pro-forms that recover their meaning anaphorically through the same mechanisms that support non-elliptical forms of discourse reference. Given this theoretical landscape, I argued that examining cases of mismatch between the elided material and its antecedent is informative for theorists from both camps: for IDENTITY theorists, mismatch cases constrain viable definitions of IDENTITY by providing potential counterexamples; and for referential theorists, they challenge our understanding of the mechanisms that enable inferential reference resolution. In line with this research program, Chapters 3 and 4 considered two types of mismatch cases under VP-ellipsis: voice mismatches in the context of the Recycling Hypothesis; and extreme cases of mismatch enabled by the presence of indirect speech acts. Chapter 5 then considered sluicing under argument-structure mismatches and tough movement, demonstrating

that sluicing, too, can be acceptable despite the presence of mismatches between the elided material and its antecedent.

This chapter will raise further challenges for IDENTITY theories of sluicing by considering novel types of mismatches that have not been discussed in the literature. Specifically, I will show that sluicing allows for similar kinds of inferential interpretations as the ones discussed in the context of VP-ellipsis in Chapter 4, and that such examples violate both semantic and syntactic notions of IDENTITY and are likewise problematic for hybrid proposals. I then go on to discuss sluices with nominal antecedents, which further challenge the assumption that sluicing is governed by an IDENTITY condition, and explore a possible explanation of these novel facts from a referential point of view.

6.1.1 IDENTITY and the No New Words constraint

Recall from Chapter 2 that purely semantic IDENTITY conditions, such as Merchant's (2001) e-GIVENness condition, are not sufficient for capturing the distribution of ellipsis. A particularly influential solution to this problem, introduced by Chung (2006) and later adopted by many others (e.g., Merchant, 2008; Tanaka, 2011a; Chung, 2013; Merchant, 2013a, 2013b), was to extend semantic IDENTITY by a lexico-syntactic condition known as the No New Words constraint.[1] As the name suggests, this condition prevents the ellipsis of any lexical material that is not provided by the antecedent, which captures the impossibility of voice-mismatched sluicing (see Chapter 5) as well as cases involving stranded prepositions with no correlate in the antecedent:

(168) a. #Jackson is jealous, but we don't know who ~~he is jealous of~~.

b. Jackson is jealous, but we don't know of who ~~he is jealous~~.

c. #Susan said she was afraid, but she didn't say what ~~she is afraid of~~.

[1] Recall also that Rudin's (2019) purely syntactic IDENTITY proposal also derives the No New Words constraint as well as some form of semantic IDENTITY, even though he doesn't explicitly reference either.

d. Susan said she was afraid, but she didn't say of what ~~she is afraid~~.

In order to prevent the No New Words constraint from being overly restrictive, Chung (2006) had to carve out several exceptions that are also echoed in Rudin (2019), illustrated in (169).

(169) a. John is eating, but I can't see what$_i$ ~~he is eating t_i~~. = (25)

b. Susan hates the Thompsons and the detectives wanted to know whether [the Thompsons]$_1$ know why ~~Sue hates them$_1$~~.

(adapted from Merchant, 1999, ex. 16)

c. [Which person]$_1$ will win the next election and by what margin ~~will they$_1$ win it~~?

(Ginzburg, 1992, p. 302a)

First, the No New Words condition must ignore syntactic traces in sprouting cases in which the sluicing remnants have no overt correlate in the antecedent (Merchant, 2013a), as in (169a). Secondly, cases involving "vehicle change" (Fiengo & May, 1994) between pronouns and their antecedents, as in (169b), have to be explicitly exempted from the No New Words constraint as well.[2] Finally, pronouns must also be exempted from the No New Words constraints in cases like (169c), even though they are not co-referential with any element in the antecedent. This exception differs from the first two in that these lexical mismatches not only violate the No New Words constraint, but also e-GIVENness: since the elided pronoun receives a non-interrogative interpretation whereas its correlate in the antecedent has an interrogative meaning, the two cannot be semantically identical (Rudin, 2019).

In this chapter, we will consider additional challenges to hybrid IDENTITY theories of

[2]This is particularly significant considering the fact that Merchant's e-GIVENness was explicitly motivated by the desire to avoid such an exception: "To pursue a theory of [syntactic IDENTITY] while considering the cases of 'vehicle change' to have been sufficiently dealt with simply by naming them is to confuse the diagnosis with the cure" (Merchant, 2001, p. 25). Thus, the return to a (partially) syntactic IDENTITY condition undermines one of the key motivations for e-GIVENness.

ellipsis that invoke a lexico-syntactic condition like the No New Words constraint. Like (169c), many of the cases we will examine violate both the No New Words constraint as well as semantic IDENTITY, but unlike "vehicle change" mismatches, the cases we will consider do not lend themselves to generalized exceptions. Consider the examples in (170), which are representative of the larger set of sluices that feature in the experiments reported in this chapter.

(170) a. A: Can I borrow your textbook?
B: Which textbook (do you have in mind; do you want/need to borrow; ...)?
b. A: Did you not tell your friends about the game today?
B: I did, but I forgot to tell them where (it would be; it would take place; ...).

In (170a), B responds to A's request to borrow their textbook with the sluiced clarification question *Which textbook?*, which is readily interpreted as *Which textbook do you have in mind?* or *Which textbook do you need?*. If the ellipsis was licensed under IDENTITY with the antecedent clause, however, the sluice would have to be interpreted as *Which textbook can you borrow?*, which is clearly not the interpretation it receives. As we will see, experimental participants are happy to interpret those sluices inferentially in a way that goes beyond the antecedent-provided meaning (i.e., deriving their interpretations under IDENTITY would violate both e-GIVENness as well as the No New Words constraint) and they nonetheless rate these utterances as highly acceptable. In (170b), the antecedent is a single noun phrase, rather than a full clause: *game*. Since the remnant wh-phrase in the ellipsis clause must compose with a proposition in order to yield a full question meaning, however, the ellipsis must be resolved to a full proposition, not just the meaning of a noun phrase. Therefore, the interpretation of such nominal-antecedent sluices will inevitably have to be inferential to some extent; in this case, the sluiced question may end up meaning something like *where the game would be* or *where the game would take place*.

Our investigation will proceed as follows. Experiment 1 will seek to verify the intuitive

judgments about preposition-stranding cases that motivated the No New Words constraint, exemplified in (168), and compare these examples to a novel set of sluices that are modeled after (170a) and prone to violating both e-GIVENness and lexico-syntactic IDENTITY. As we did in the context of VP-ellipsis in Chapter 4, we ask participants to both judge and paraphrase the ellipsis clause, which allows us to estimate the number of elided lexical items that are not provided by the antecedent and thus must be exempted from the No New Words constraint in order classify those cases as grammatical under IDENTITY. Experiments 2-4 will then zero in on a subset of the examples in Experiment 1, which involve nominal antecedents. In particular, we will test whether the predictability of the to-be-sluiced questions affects the acceptability of sluicing—a hypothesis that is echoed in several recent theories of sluicing.

6.2 Experiment 1: sluices with inferred readings

The purpose of this experiment was twofold: to experimentally verify the armchair judgments reported in the literature that motivated the No New Words constraint; and to test it against a novel set of sentences that may receive interpretations that involve "New Words" and nonetheless appear to be relatively acceptable. In Chapter 4, we have seen a class of inferentially resolved cases of VP-ellipsis, and the present experiment aims to establish the existence of such cases for sluicing by taking a similar approach. As before, we examine short dialogues and employ an experimental paradigm that both elicits acceptability judgments and—in the second part of the experiment—asks participants to paraphrase the ellipsis site in order to determine the meaning it acquires.

6.2.1 Methods

Materials

We constructed a set of 50 experimental items: 10 minimal pairs that I will be referring to as "CLASSIC" because they were modeled after sentences from the literature that motivated the No New Words constraint; 20 novel items that I will be referring to as "INFERENCE" items because they were designed to facilitate inferential ellipsis resolution; and 20 filler sentences that involved sluicing did not incentivize inferential readings or involve preposition stranding. Half of the fillers were acceptable while the other half were unacceptable due to a variety of grammatical violations.

All experimental and filler items consisted of a 2-turn dialogue in which the first utterance provided the antecedent and the second utterance contained a sluiced question. The CLASSIC items consisted of minimal pairs like (171), in which the ellipsis clause was paired with an antecedent clause that either did or didn't include a correlate to the elided preposition. The INFERENCE items were designed with the goal of eliciting inferential readings, as exemplified in (172).

(171) a. Claire: Joe is jealous of someone.
Jessica: Do we know who? [acceptable CLASSIC item]

b. Claire: Joe is jealous.
Jessica: Do we know who? [unacceptable CLASSIC item]

(172) Fan: Can I get a few autographs?
Manager: Sure, how many? [INFERENCE item]

It is important to note that the INFERENCE items form a heterogeneous set of examples that were carefully designed to be as acceptable as possible. Rather than uncovering a pattern that generalizes to all instances of sluicing, our goal was to pursue a *proof of concept* that addresses the

following question: is it possible to construct contexts in which it is felicitous to resolve sluiced questions inferentially rather than relying solely on their antecedent for their interpretation?

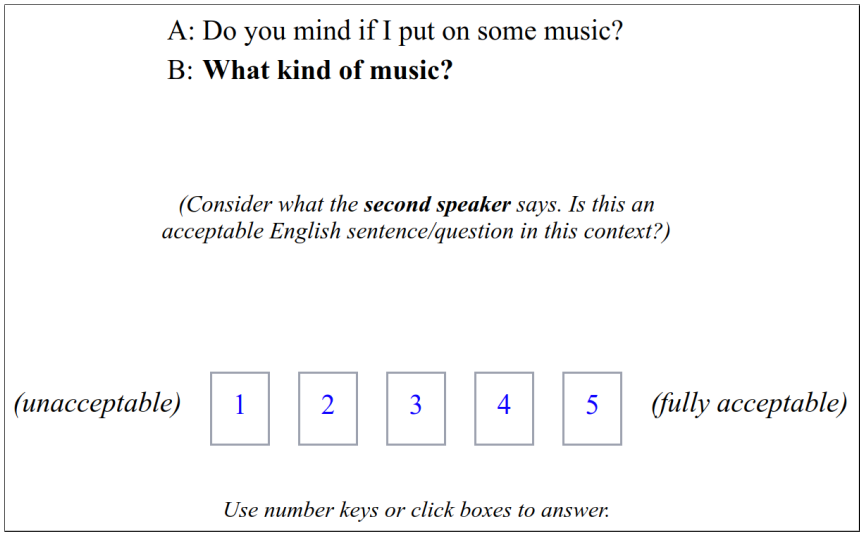
Procedure

As in the experiment described in Chapter 4, participants performed two tasks in separate blocks. First, they were instructed to evaluate the acceptability of the elliptical utterance ("what the second speaker says") in the context of the first utterance on a scale from 1 ("unacceptable") through 5 ("fully acceptable"). In the second block, participants were asked to paraphrase the utterances they had previously seen in the context of the acceptability judgment task. Specifically, they were asked "What exactly does the second speaker mean?" and told: "While you may use your own words as you type your response, please try to express as precisely as possible what the speaker meant to say." The instructions for both tasks were also included in abbreviated form on every trial to ensure that participants knew what they were asked to do throughout the experiment. The screenshots in Figure 6.1 show an example trial involving each task from the perspective of the participants.

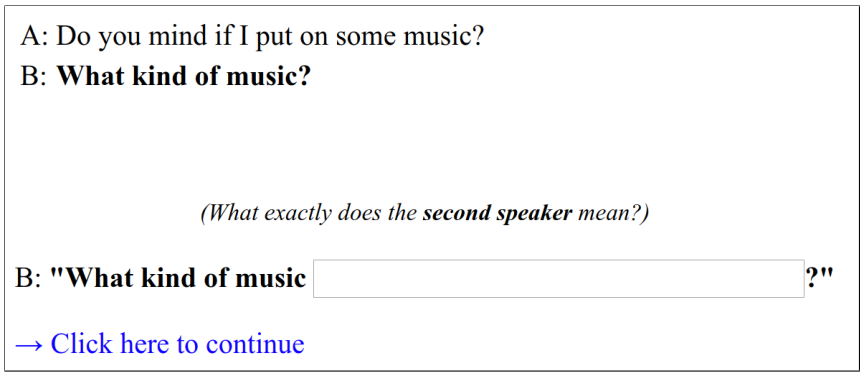
Analyses & predictions

The acceptability results were analyzed as in previous chapters: we fit a hierarchical ("mixed-effects") ordinal regression model using the `brms` R package (Bürkner, 2017, 2018). Since we are interested in the differences between the three types of items that were included in the experiment (ACCEPTABLE CLASSIC, UNACCEPTABLE CLASSIC, and INFERENCE items), we included item type as a treatment-coded population-level predictor. As usual, we also added crossed by-subject and by-item group-level intercepts as well as slopes for each population-level predictor. We predicted the novel set of INFERENCE items to pattern with the acceptable variants of the CLASSIC items rather than their unacceptable counterparts.

The second dependent variable, the number of New Words contained in each paraphrase,



(a) Acceptability rating task



(b) Paraphrase task

Figure 6.1: Screenshots of the acceptability rating task (top) and the paraphrase task (bottom).

was calculated by comparing the paraphrase to the antecedent clause and counting how many of its words did not occur anywhere in the antecedent clause. This was done in a way that ignored obvious spelling errors (e.g. "unload" vs. "unloadt"), differences in upper/lower case, morphological differences (e.g. "say" vs. "said"), contractions ("he's" vs. "he was"), and other differences in spelling with no bearing on the "No New Words" constraint (e.g. "interesting-looking" vs. "interesting looking"). We further removed 71 responses from participants who used the text box to paraphrase the entire ellipsis clause, including the remnants, rather than merely filling in the "missing" material. Finally, we never counted "do it" or "do that" as New

Words. Once the number of New Words was determined for each paraphrase, we analyzed these measures by fitting a hierarchical poisson regression model with maximal random effects using `brms` (Bürkner, 2017, 2018) described in more detail below.

We expected the CLASSIC items to be bimodally distributed in terms of both acceptability and the number of New Words: by design, the acceptable variant of each minimal pair includes the target preposition in the antecedent clause, which should lead to fewer New Words in participants' paraphrases. Since the Number of New Words measure is a novel measure, this prediction will serve to verify the task and establish a baseline against which the other types of items can be compared. As for the INFERENCE items, we expect a relatively large number of New Words if they do, in fact, succeed in eliciting inferential interpretations. If the No New Words constraint is the correct explanation for the unacceptability of the lower-bound CLASSIC items, and if the INFERENCE items do involve New Words as well, we should find them to be unacceptable as well.

6.2.2 Results

Data exclusion

We recruited 28 participants via Amazon.com's Mechanical Turk. One of those participants was excluded for self-identifying as a non-native speaker of English. The data from the remaining 27 participants was further filtered to exclude 9 individual trials on which the response was submitted in less than 1000 ms (we assume that reading and judging the materials carefully would take at least 1 second), leaving us with a total of 1341 individual observations.

Number of "New Words"

We calculated the number of words in participants' paraphrases that did not occur in the antecedent clause following the procedure described above. Figure 6.2 shows the average number

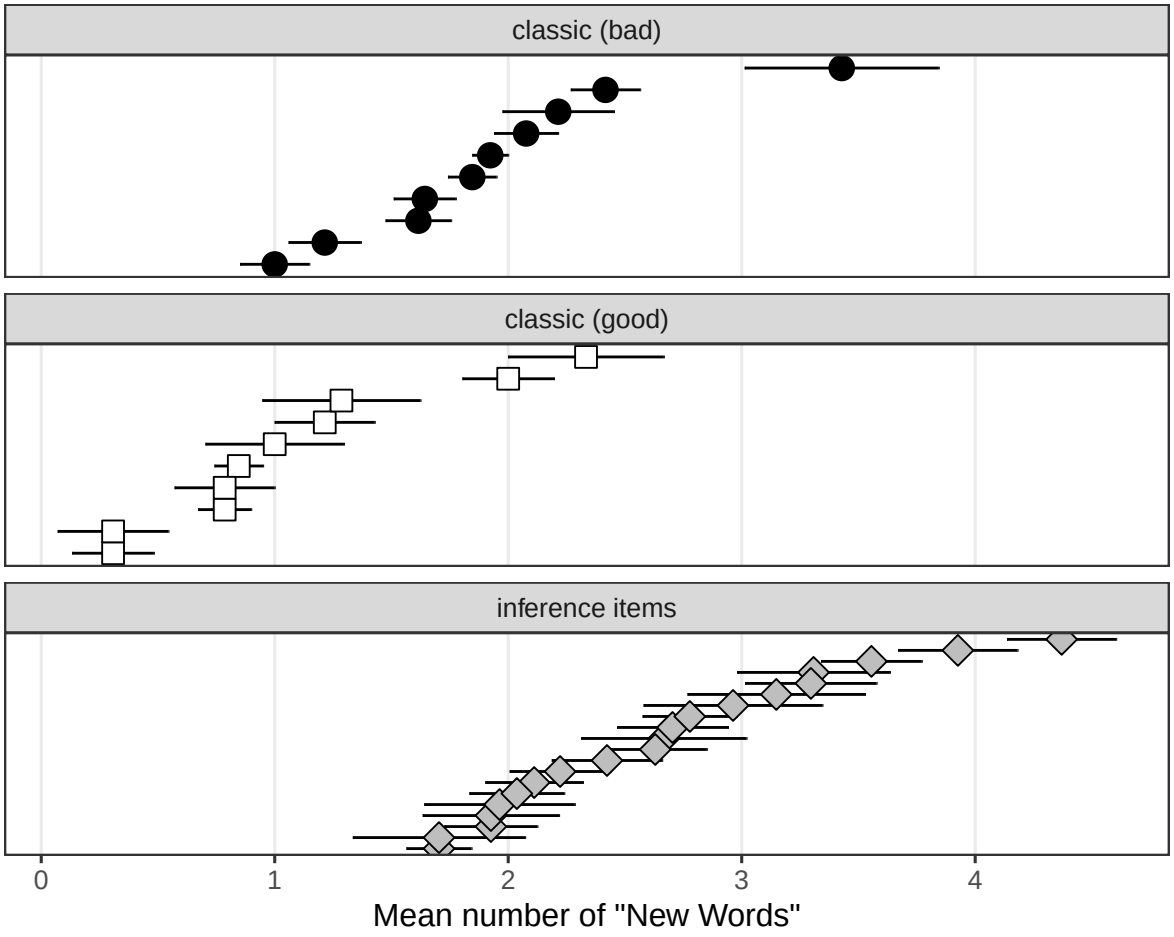


Figure 6.2: Mean number of New Words in participants' paraphrases of the ellipsis site, grouped by item type: good and bad variants of CLASSIC minimal pairs from the literature (top and middle); and INFERENCE items (bottom). Error bars represent Standard Errors.

of words per item.

We analyzed the results using a multi-level Poisson regression with item type as a treatment-coded population-level effect (ACCEPTABLE CLASSIC, UNACCEPTABLE CLASSIC, and INFERENCE items) and all by-item and by-participant group-level ("random") effects justified by the design (Barr et al., 2013). As expected, the acceptable variants of the CLASSIC minimal pairs involved significantly fewer New Words than their unacceptable counterparts ($\Delta = -0.86$, $CI(\Delta) = [-1.2, -0.51]$, $P(\Delta > 0) = 1$; average difference: 0.86 words). This is an important result because it suggests that the paraphrase task is capable of revealing the difference we are interested

in, even though the absolute numbers may not be interpretable directly: while even paraphrases of the acceptable variants contained on average 1.07 words not provided by the antecedent, their unacceptable counterparts involved on average an additional 0.86 such words. That being said, our key prediction was that the INFERENCE items would involve a large number of New Words and that is exactly what we found: items in this group involved on average 1.93 New Words, significantly more than both the acceptable CLASSIC items ($\Delta = 0.73$, $CI(\Delta) = [0.47, 1]$, $P(\Delta > 0) = 1$) as well as their unacceptable counterparts ($\Delta = 1.59$, $CI(\Delta) = [1.32, 1.86]$, $P(\Delta > 0) = 1$).

Acceptability

Figure 6.3 shows the average acceptability of experimental items across categories. The "classic" items (top two facets) exhibit the expected bimodal distribution. The INFERENCE items (bottom) pattern with the acceptable CLASSIC items, even though paraphrasing their meaning resulted in a relatively large number of New Words. We modeled the results in a multi-level ordered-category model with item type as a treatment-coded population-level effects and the maximal random-effect structure, consisting of by-item and by-subject group-level intercepts and slopes for item type (Barr et al., 2013). As expected, ACCEPTABLE CLASSIC items were significantly more acceptable than UNACCEPTABLE CLASSIC items ($\Delta = 2.25$, $CI(\Delta) = [1.71, 2.81]$, $P(\Delta > 0) = 1$), and so were the novel inference items ($\Delta = 1.99$, $CI(\Delta) = [1.21, 2.77]$, $P(\Delta > 0) = 1$). Finally, inference items did not differ significantly from the ACCEPTABLE CLASSIC ones ($\Delta = -0.26$, $CI(\Delta) = [-0.99, 0.43]$, $P(\Delta < 0) = 0.77$), which is consistent with our expectation that they would pattern together in terms of acceptability.

Acceptability as a function of the number of New Words

The scatter plot in Figure 6.4 shows the relationship between the mean number of New Words and the average acceptability associated with each item. The negative slopes on the lines that connect the two variants of the CLASSIC minimal pairs indicate that with respect to those

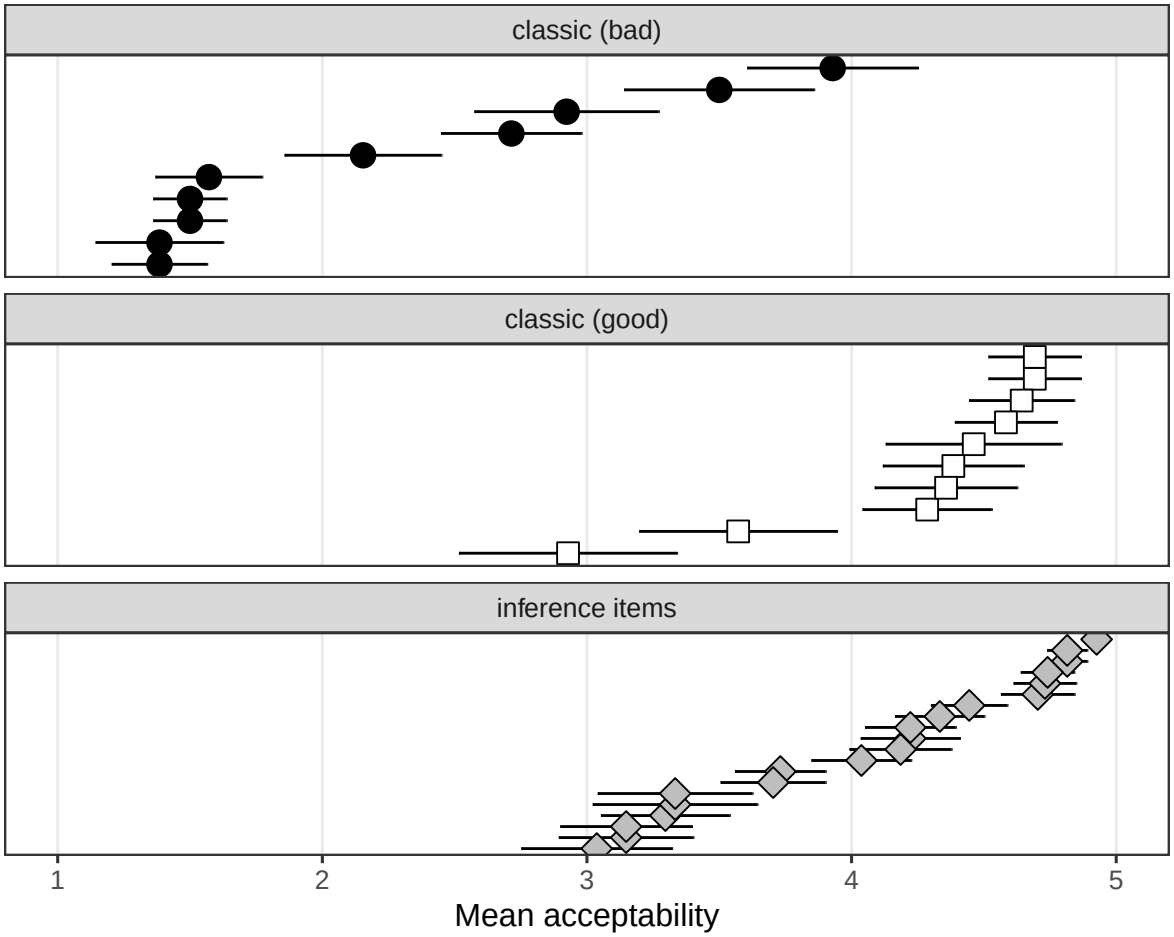


Figure 6.3: Acceptability results from Expt 1, grouped by item type: good and bad variants of CLASSIC No New Words minimal pairs from the literature (top and middle); and novel INFERENCE items (bottom). Error bars represent Standard Errors.

items, an increase in the number of New Words contained in paraphrases of the ellipsis was associated with a dramatic reduction in acceptability. However, this plot also reveals that this relationship does not hold with respect to the INFERENCE items: they involve an even greater number of "New Words" but were nonetheless rated as relatively acceptable.

To quantify these observations statistically, we conducted a multi-level analysis with raw ratings as the dependent variable and the number of new words as a continuous predictor along with a 2-way ITEM TYPE factor distinguishing the classic items from the inference items and the interaction between them. As always, we added all group-level intercepts and slopes for items and

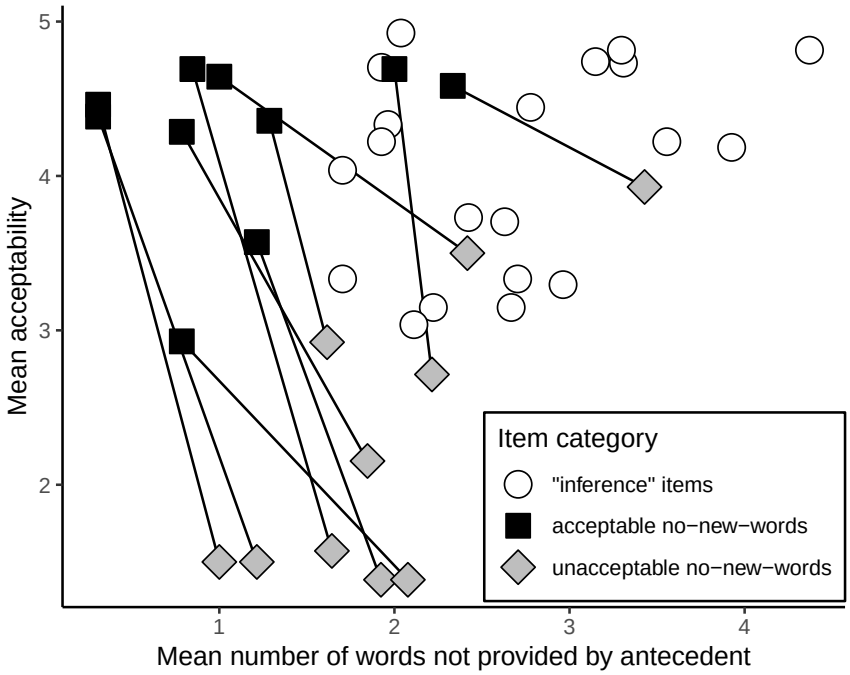


Figure 6.4: Scatter plot showing results from Experiment 1. The mean acceptability of each item (y) is plotted in terms of the average number of New Words it produced in the paraphrase task (x). Lines connect variants of the same CLASSIC minimal pair.

participants that were justified by the design. The results confirm the pattern shown in Figure 6.4: within the CLASSIC items, increasing numbers of New Words were associated with a significant drop in acceptability ($\Delta = -0.65$, $CI(\Delta) = [-0.84, -0.48]$, $P(\Delta < 0) = 1$), whereas the opposite was true for INFERENCE items ($\Delta = 0.72$, $CI(\Delta) = [0.52, 0.95]$, $P(\Delta > 0) = 1$).

6.2.3 Discussion

The goal of Experiment 1 was to test the No New Words generalization against (a) preposition-stranding examples sourced from the literature, and (b) a novel set of items that were specifically designed to encourage comprehenders to draw inferences. The CLASSIC items behaved exactly as the No New Words generalization would predict: within each minimal pair, the variant that led to an increased number of New Words in participants' paraphrases of the ellipsis site was also the variant that was associated with a dramatic decrease in acceptability.

However, the generalization failed to predict the behavior of the novel set of INFERENCE items that were specifically designed to facilitate inferential ellipsis resolution. We therefore conclude that whatever is causing CLASSIC items to become unacceptable, it cannot be a general ban against eliding words that are not provided by the antecedent. In fact, as I argued at length in Chapter 2, the possibility of inferential ellipsis resolution represents a challenge for IDENTITY theories of sluicing more broadly: all of the INFERENCE items in this experiment violate IDENTITY in one way or another and should consequently be ungrammatical. The fact that most of them are highly acceptable, then, is inconsistent with the IDENTITY requirement.

The results from Experiment 1 raise an important question that we will address in the remainder of this chapter: What supports the inferential resolution of sluicing? In the following section, we take a closer look at a subset of the INFERENCE items from Experiment 1 and develop a working hypothesis that is rooted in referential theories of sluicing, and the remainder of this chapter is concerned with testing this hypothesis across several experiments.

6.3 Sluicing and question predictability

6.3.1 Sluicing with nominal antecedents

One of the most highly rated examples in Experiment 1 is shown in (173), along with the interpretation most participants adopted (in brackets).

(173) A: Did you not tell your friends about the game today?
B: I did, but I forgot to tell them where (the game is).

The overwhelming majority of participants in Experiment 1 interpreted the sluiced question as "where the game is/was." That interpretation, and the fact that the sluice is perfectly acceptable, would be entirely unsurprising if the context contained an antecedent clause that denotes this

proposition, as in (174).

(174) A: Did you not tell your friends that the game is today?
B: I did, but I forgot to tell them where (the game is).

But unlike (174), the context in (173) does not provide a clause that corresponds directly to the meaning of the ellipsis site. Instead the noun phrase *game* appears to serve as the antecedent for the ellipsis clause, which then simply goes on to ask a *where* question about the game.

The fact that a mere NP can serve as the antecedent for clausal ellipsis raises a serious challenge for IDENTITY theories of sluicing since noun phrases and clauses are fundamentally distinct linguistic objects, both semantically and syntactically. Despite their theoretical significance, however, nominal-antecedent sluices have received almost no attention in the literature (but see Beecher, 2007, ex. 8), which is particularly striking given the fact that analogous facts have received ample attention in the literature on VP-ellipsis (Hardt, 1993; Kehler, 1993b; Johnson, 2001; Merchant, 2013a; Miller & Hemforth, 2014) Consider the following examples, repeated from (36) in Chapter 2:

(175) a. The letter deserves a response, but before you do (respond), ...
(Kehler, 1993b, ex. 20)

b. Meanwhile, they sense a drop in visitors to the city. Those who do (visit the city), they say, are not taking cabs. (Kehler, 1993b, ex. 21)

c. Harry used to be a great speaker, but he can't (speak) anymore, because he has lost his voice. (Hardt, 1993, ex. 114)

Each of these examples is perfectly felicitous despite the fact that the elided VP and its nominal antecedent are distinct both semantically and syntactically. The standard response from IDENTITY theorists has been to argue that each of these nominal antecedents underlyingly contains a VP that

serves as an IDENTITY-preserving antecedent (Fu et al., 2001; Johnson, 2001; Merchant, 2013a). In Chapter 2 I argued that this line of reasoning is ultimately unsatisfying with respect to VP-ellipsis, but it is even more problematic if applied to sluicing: While NPs like *response*, *visitors* and *speaker* can be reasonably analyzed as deverbal, i.e. as deriving from the corresponding verbs, the NP *game* is not "de-clausal" in the same way and there are no independent reasons for analyzing it as underlyingly containing a full clause to serve as the antecedent for the sluice.[3]

To make matters even more complicated, any analytic strategy that categorically permits nominal-antecedent sluices like (173) is challenged by the fact that the exact same context that enables the *where* sluice does not appear to be adequate for sluicing a question with a different remnant wh-phrase:

(176) A: Did you not tell your friends about the game today?
B: I did, but I forgot to tell them how long #(the game is).

We are thus in a position where the ellipsis of the same clause (*the game is t*) in one and the same context with the same nominal antecedent is felicitous for some remnants, but infelicitous for others.[4]

This state of affairs is similar to the pattern we discovered in Chapter 5, whereby *when* and *where* sluices were significantly less acceptable than *how* variants of the same items, even though everything except for the remnant wh-phrase itself—including the antecedent and the elided material—were held constant in this comparison:

(177) Brake fluid is easy to replace if you know...
a. ...how (to replace it).
b. ...when #(to replace it).

[3] See Hardt, Anand, and McCloskey (2020) for a proposal that nonetheless pursues this type of strategy.

[4] An open question at this point is whether the unelided variant of (176) is acceptable or if it is just as marked as the sluiced variant. We will address this question in Section 6.6 below.

c. ...where #(to replace it).

Based on an item-by-item analysis of this penalty for *when* and *where* questions relative to the *how* baseline, I suggested that it may be driven, or at least modulated, by the plausibility of the sluiced question given its context (see Section 5.2.2). In the remainder of the present chapter, I will explore to what extent nominal-antecedent sluices like (173) and (176) are amenable to a similar analysis. As we will see in the next section, this hypothesis is consistent with a line of theories that link the use of ellipsis to the availability of a relevant Question Under Discussion (QUD; Roberts, 1998; Ginzburg & Sag, 2000; Roberts, 2012) in the context of utterance.

6.3.2 Sluicing and the Question Under Discussion (QUD)

I argued in Chapter 2 that various patterns associated with VP-ellipsis can be analyzed in terms of the extent to which the context raises a relevant Question Under Discussion (Kertz, 2013; Miller & Pullum, 2013; Miller & Hemforth, 2014; Kehler, 2016). Most relevant for our present purposes is Miller and Hemforth's (2014) work on VP-ellipsis with nominal antecedents. While the mismatch between the elided material and its antecedent in such cases violates the IDENTITY condition (both semantic and syntactic variants), Miller and Hemforth (2014) propose that the acceptability of nominal-antecedent VP-ellipsis instead depends on the extent to which the antecedent NP raises a concealed question, as indicated in square brackets in following examples, repeated from (84).

(178) a. Mubarak's survival [= whether he will survive] is impossible to predict and, even if he does (survive), his plan to make his son his heir apparent is now in serious jeopardy. (Miller & Hemforth, 2014, ex. 1)

b. The integrity of the Senate depends on her participation [= whether she participates]. If she does (participate), . . . (Miller & Hemforth, 2014, ex. 10a)

c. The release of this information on the user's part depends on his consent [= whether he consents]. If he does (consent), . . . (Miller & Hemforth, 2014, ex. 10c)

According to Miller and Hemforth (2014), VP-ellipsis in these examples is enabled by the fact that the processing of the antecedent NP requires comprehenders to access the meaning of the concealed question, which makes it accessible for subsequent ellipsis.

Importantly, however, it is not the antecedent NP itself that enables VP-ellipsis, as demonstrated by the following minimal pair:

(179) a. The letter deserves a response, but before you do (respond)... (Kehler, 1993a)

b. The letter contains a response, but before you do #(respond)...

While the context in (179a) succeeds in raising the QUD that is to be addressed by the ellipsis clause, (179b) does not, even though it contains the same antecedent NP *response*. It thus appears that the antecedent alone does not determine whether or not ellipsis is felicitous. Rather, it must conspire with the other information provided by the context in order to raise the relevant QUD.

QUD-based approaches to ellipsis have not been confined to VP-ellipsis (Ginzburg & Sag, 2000; AnderBois, 2010, 2014; Barros, 2014). For example, AnderBois's (2014) influential theory of sluicing posits a QUD-based well-formedness condition according to which sluicing is felicitous only if the sluiced question corresponds to a salient QUD. In order for this sort of theory to avoid circularity and have any explanatory value, there has to be an independently motivated measure of QUD availability that is grounded outside the theory of ellipsis. AnderBois addresses this issue by drawing on concepts from Inquisitive Semantics: he credits the inner antecedent with the primary responsibility for raising the QUD that enables sluicing. For example, the inquisitive semantic contribution of the (wide-scope) indefinite *something* in (180) is responsible for raising the QUD *What did Janine hear?* and thus allows that question to be sluiced.

(180) Janine heard something, but she doesn't want to say what (she heard).

While this approach has the advantage of defining precisely how QUDs are raised, it is overly restrictive in a way that undermines the generality of the proposal. For example, it is widely assumed that QUDs can be raised in a variety of ways and not just through the presence of inquisitive linguistic elements like indefinites and disjunctions (Roberts, 1998, 2012). Indeed, AnderBois' approach is insufficient for addressing the examples we are interested in here, repeated below.

(181) A: Did you not tell your friends about your game today?

B: I did, but forgot to tell them...

a. ...where (the game is).

b. ...how long #(the game is).

Since there is no inner antecedent, neither the *where* question nor the *how long* question correspond to a salient QUD under his account, making it impossible to explain the contrast between them.[5]

Instead of relying on the presence of inquisitive elements, we will therefore take a different approach towards defining QUD availability. Specifically, we will not only conduct an acceptability judgment study, but pair it with a separate experiment in which participants are presented with the exact same prompts and perform a forced-choice passage completion task. This will allow us to estimate comprehenders' beliefs about the likelihood of each of the sluiced questions independently of sluicing, which we can then correlate with the acceptability of sluicing those questions.[6]

[5] AnderBois (2014) does suggest that cases without inquisitive inner antecedents may instead rely on bridging inferences, but he does not offer a theory of bridging that would drive a wedge between the examples we are considering here.

[6] This strategy is similar to Miller and Hemforth's (2014) approach to nominal-antecedent VP-ellipsis: they experimentally measured the extent to which the antecedent NP gave rise to a concealed polar question and correlated this score with the acceptability of using VP-ellipsis.

6.4 Experiment 2: nominal-antecedent sluices

The goal of this experiment was to test the acceptability of sluices with nominal antecedents and to lay the groundwork for testing whether the acceptability of such sluices depends on the availability of the relevant Question Under Discussion (QUD). To that end, we constructed a set of experimental items in a way that allowed us to hold the context of utterance constant while varying the sluiced question and how predictable that question is given the context. In the next experiment, we will then use the exact same contexts to estimate how predictable each question is and test whether that predictability score is correlated with the acceptability of sluicing.

6.4.1 Methods

Materials

We constructed 30 nominal-antecedent sluices that were similar to the example in (173) in that they all involved a nominal antecedent. In particular, we constructed six contexts and substituted in five different sluicing remnants, as shown in (182).

(182) When you call 911 about an emergency, the first thing they ask is...

a. ...where.

b. ...why.

c. ...what exactly.

d. ...who.

e. ...when.

The goal behind this item writing procedure was to expose variability in acceptability: while (182a) seems reasonably acceptable, (182b-c) may be somewhat degraded and (182d-e) downright unacceptable. Note, however, that the entire context, including the (nominal) antecedent as well

as the sluice-embedding clause, was held constant while only the sluicing remnants were varied across item variants.

Participants & procedure

We recruited 63 participants via Amazon.com's Mechanical Turk, all of whom self-identified as native speakers of English. Each participant saw six of the 30 experimental items (one from each context) along with 12 filler items, half of which were designed to be acceptable while the other half was meant to establish the lower end of the acceptability scale.

6.4.2 Results

We excluded data from trials that lasted less than 1,000 ms under the assumption that it would take at least that much time to read each sentence and submit an acceptability judgment, leaving us with a total of 1012 individual observations to analyze. The results, which are summarized in Figure 6.5, reveal a large amount of variation: some nominal-antecedent sluices were rated extremely highly, some were judged to be highly unacceptable, and the remaining items covered the entire range of possible acceptability ratings in-between.

6.4.3 Discussion

We designed the stimuli for this experiment with the explicit goal of exposing variability that we can then try to explain in terms of QUD predictability. The results reveal that the item creation procedure was successful in this regard: the experimental items covered the entire range of possible acceptability judgments between the bounds established by the filler items.

These results speak to and represent a challenge for all existing theories of sluicing. For IDENTITY theories, they are problematic because all nominal-antecedent sluices are predicted to be categorically ungrammatical, whereas we found some of them to be highly acceptable.

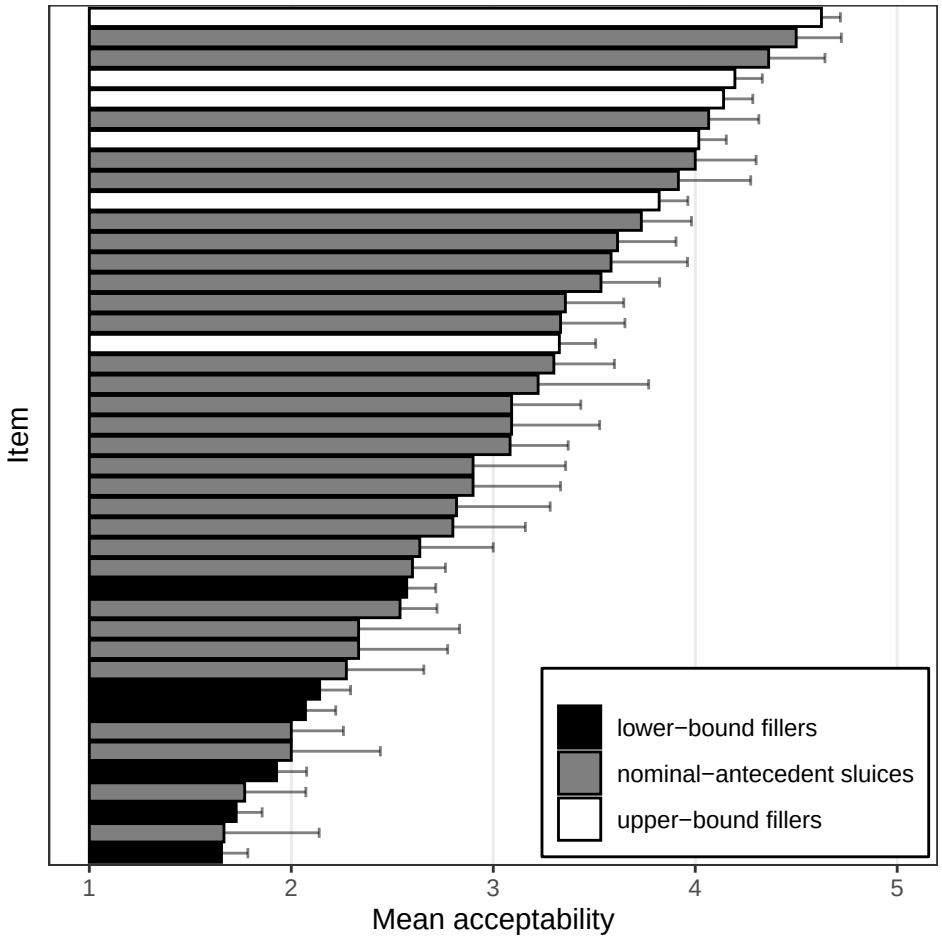


Figure 6.5: Results from Experiment 2. Bars show mean by-item acceptability ratings of acceptable fillers (white), unacceptable fillers (black), and nominal-antecedent sluices (grey). Error bars indicate Standard Errors.

Furthermore, these results cannot be explained by "fine-tuning" the definition of IDENTITY (see Section 2.1.1): if we allow sluices with nominal antecedents in order to explain the fact that some of them are highly acceptable, we face the opposite challenge of explaining why others are completely unacceptable. In fact, the extreme gradience that is evident in the results call into question whether binary classification can be successful at all.

The results are also informative for referential theories of sluicing. According to such theories, contexts with nominal antecedents can give rise to felicitous sluicing as long as the intended referent is inferrable and reasonably salient from the context. Since these theoretical

constructs are themselves gradient, the variability in the results can potentially be explained under referential theories, and the goal of the next experiments is to attempt to do so in terms of QUD predictability.

6.5 Experiment 3: QUD predictability

6.5.1 Introduction

The purpose of Experiment 3 was to directly test the hypothesis that the acceptability of nominal-antecedent sluices can in part be explained as a function of the predictability of the sluiced question. To that end, we need to quantify the question predictability associated with each item, and we will do so in a forced-choice passage-completion experiment.

6.5.2 Methods

Materials, participants, and procedure

The experimental materials were identical to those in Experiment 2: 30 nominal-antecedent sluices, consisting of six minimal sets with five variants each that differed only with respect to the sluicing remnants. 54 participants were recruited via Amazon.com's Mechanical Turk and presented with six of the 30 nominal-antecedent sluices, one from each context. Unlike Experiment 1, however, the items were truncated before the remnant wh phrase, followed by 5 possible continuations presented in a forced-choice format, as illustrated in (183).

(183) When you call 911 about an emergency, the first thing they ask is...

a. ...where you are located.

b. ...why you are calling.

c. ...what exactly the emergency is.

d. ...who they are speaking to.

e. ...when the emergency happened.

Each continuations expressed the meaning of one of the five different sluiced questions that occurred together with this context in Experiment 2, and were determined in a separate norming experiment described below. Participants were instructed to choose the most likely continuation in this context.

Norming experiment: estimating the meaning of the sluiced questions

In order to estimate the meaning of each of the sluiced questions, a separate group of 46 participants was recruited and presented with a paraphrase task that was analogous to the one in Experiment 1 (see bottom panel in Figure 6.1).[7] Specifically, they read each item in its entirety (the context as well as the sluiced question) and were then asked to paraphrase the ellipsis site using their own words. The within-item modal response for each of the 30 experimental items was selected and served as an answer choice in the forced-choice passage completion task in Experiment 3.

Predictions

We expect the predictability of the sluiced questions in our materials to vary both within contexts and across. If felicitous sluicing depends on the salience of the relevant QUD in context, we would expect the acceptability scores from Experiment 2 to be positively correlated with the question-predictability scores collected in Experiment 3.

[7] 11 participants were excluded for either self-identifying as a non-native speaker of English or for failing to perform the task correctly. For example, several participants copied a random portion of the experimental prompt into the text box, which is behavior exhibited by automated MTurk bots.

6.5.3 Results

Three participants were excluded for self-identifying as non-native speakers of English, and an additional 53 observations were excluded because they were submitted in less 5000 ms (we assume that it takes at least 5 seconds to carefully read the prompt and all answer choices), leaving us with a total of 253 individual observations from 47 participants.

The results are shown in Figure 6.6. In each context, some questions were considered more likely considerations than others. For example, given the context *Regarding Trump's impeachment, the only question is...*, the most likely continuation was the question *when it will happen*, where as the questions *who will impeach him* and *where the impeachment is likely to occur* were extremely unlikely. The key question for Experiment 3 was whether the variability in question predictability, summarized in Figure 6.6, is correlated with the acceptability of sluicing those questions, which was measured in Experiment 2.

Figure 6.7 suggests that there is indeed a positive correlation: the more likely a question was to be selected in the forced-choice passage completion task, the higher its expected acceptability rating. To confirm whether this relationship is statistically significant we conducted a hierarchical ordinal regression analysis with acceptability as a categorical outcome variable and question predictability (i.e., the within-item proportion of trials on which the relevant question was selected in the passage completion task in Experiment 3) as a population-level predictor. We additionally included by-item and by-participant group-level intercepts and slopes for question predictability. The results confirm that acceptability of sluicing does indeed increase significantly as a function of question predictability ($\Delta = 2$, $CI(\Delta) = [-0.23, 4.2]$, $P(\Delta > 0) = 0.97$).

6.5.4 Discussion

The purpose of Experiment 3 was to test whether the acceptability of sluicing questions in nominal-antecedent contexts is correlated with the predictability of those questions given

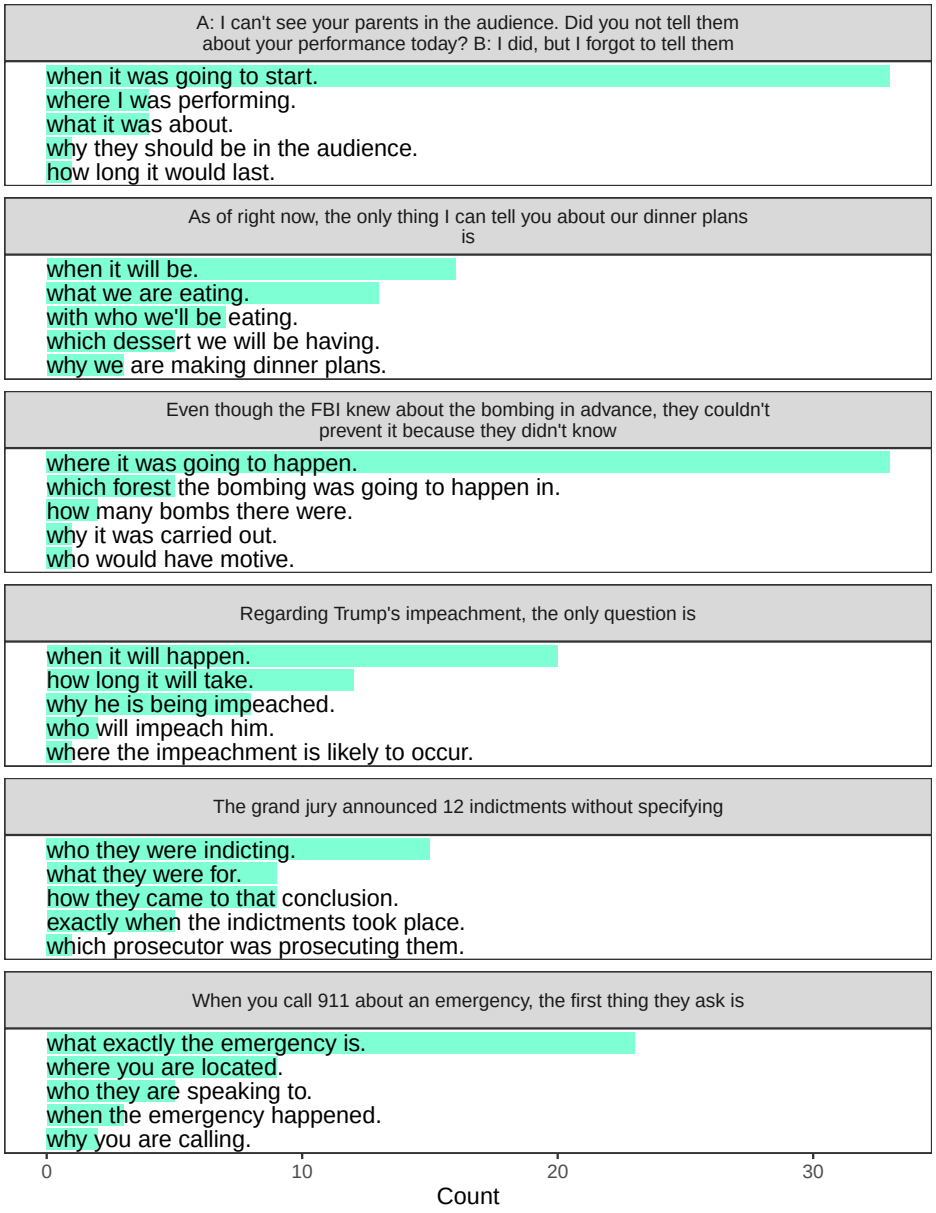


Figure 6.6: Results from Experiment 3. Bars indicate the number of times participants chose each question as the most likely continuation given the context.

the context. To that end, we operationalized question predictability in terms of a forced-choice passage completion task in which participants were presented with the same contexts that featured the sluiced questions in Experiment 2 and selected the most likely continuation. Both the predictability of the questions and the acceptability of sluicing them exhibited variability and, as

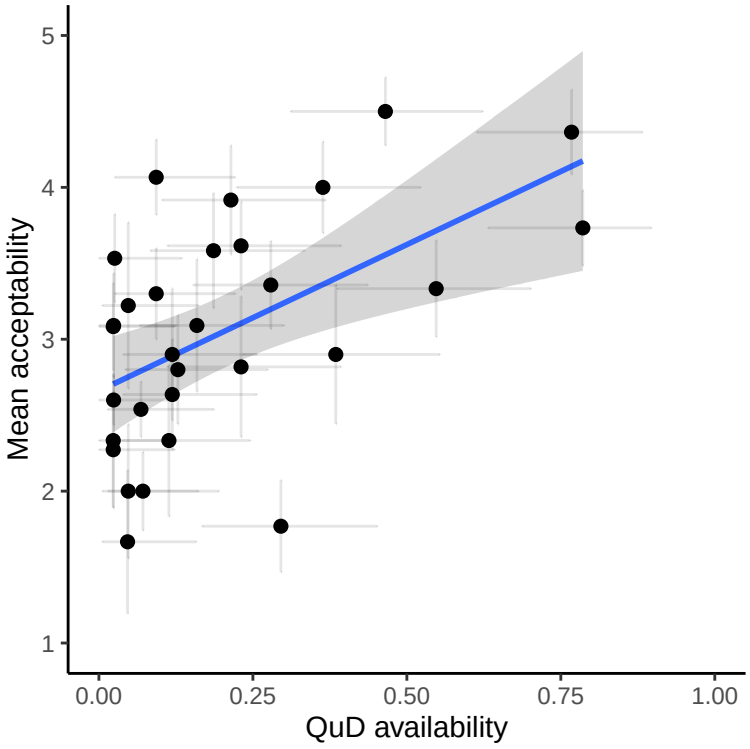


Figure 6.7: Acceptability (Experiment 2) as a function of QUD availability (Experiment 3). Vertical error bars reflect Standard Errors; horizontal error bars show 95% exact confidence intervals.

predicted, the two measures were positively correlated.

What we don't know at this point is whether question predictability is predictive of the acceptability of sluicing *per se* or whether it is more generally associated with the acceptability of the question regardless of whether it is sluiced or not. The reason for that is that Experiment 2 did not include unelided variants of the sluiced questions because it is non-trivial to determine the meaning of those questions. Since we had to establish the meaning of each question for the passage-completion paradigm in Experiment 3, however, we can now test the acceptability of the unelided variants to see whether the question predictability effect is specific to sluicing or not.

6.6 Experiment 4: acceptability including unelided variants

The goal of Experiment 4 was twofold: to replicate the findings from Experiments 2 and 3; and to test whether the same pattern holds for unelided variants of the sluiced questions. We thus replicated the acceptability judgment task from Experiment 2 and included unelided variants of each item. Those unelided variants were determined by selecting the modal response from the norming experiment we conducted in preparation for Experiment 3 in which participants paraphrased the ellipsis site.

6.6.1 Methods

Materials

The materials were identical to those used in Experiment 2 except that unelided variants (see Figure 6.6) were added, as shown in (184).

(184) When you call 911 about an emergency, the first thing they ask is...

a. ...where (you are located).

b. ...why (you are calling).

c. ...what exactly (the emergency is).

d. ...who (they are speaking to).

e. ...when (the emergency happened).

As in Experiment 2, the experimental items were interspersed with 12 filler items (2:1 ratio), half of which were acceptable and half unacceptable.

Procedure & participants

We recruited 181 participants via Amazon.com's Mechanical Turk. 19 of those participants were excluded for self-identifying as non-native speakers of English. The data from the remaining 162 participants was further filtered to exclude 612 individual trials on which the response was submitted in less than 1000 ms,[8] leaving us with a total of 2916 individual observations.

Participants read each experimental item and judged it in terms of its acceptability, using a 5-point Likert scale ranging from "unacceptable" to "fully acceptable."

Predictions

We expect to replicate the results from Experiment 2: the acceptability of sluiced questions should increase as a function of the degree to which those questions are predictable from the context. The key question is whether or not the unelided variants show the same effect. If the question predictability effect is specific to sluicing, we should expect to see a significant interaction whereby the effect is significantly less pronounced for unelided variants.

6.6.2 Results

The results are shown in Figure 6.8, which suggests that the finding from Experiment 3 is replicated fully: the acceptability of sluiced questions (left panel) appears to increase with question predictability. The novel unelided variants, however, also exhibit a positive, albeit somewhat attenuated, relationship with question predictability. Since they are overall much more acceptable than their sluiced counterparts and are pushing up against the upper bound of the acceptability scale, the attenuation of the question predictability effect may be due to a ceiling effect.

[8] As before, we assume that it is not possible to carefully read and judge the experimental items in less than a second.

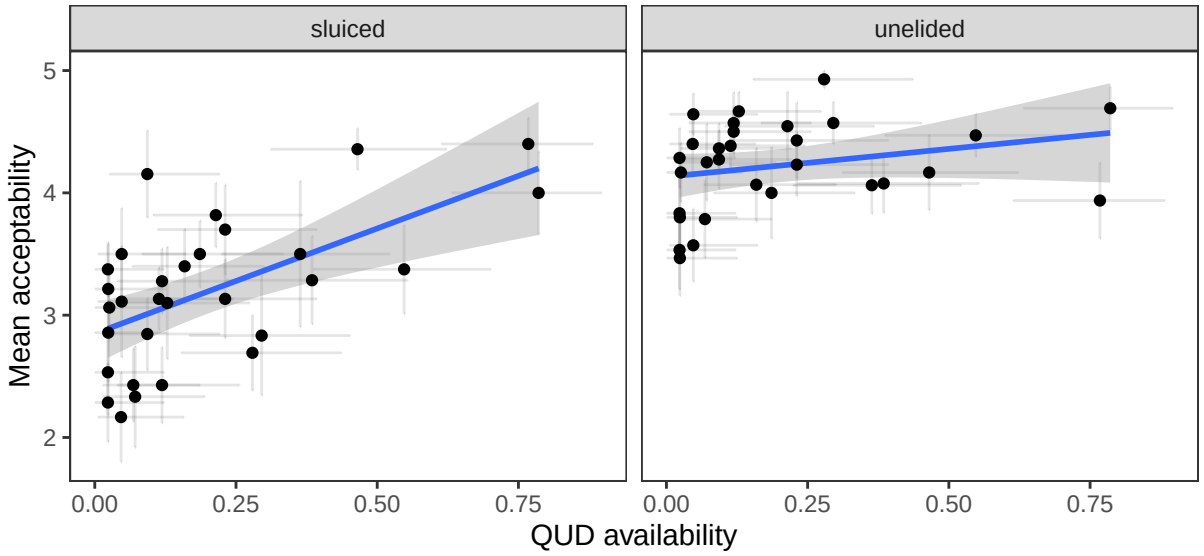


Figure 6.8: Acceptability (Experiment 4) as a function of QUD availability (Experiment 3). Vertical error bars reflect Standard Errors; horizontal error bars show 95% exact confidence intervals.

To test these patterns statistically, we fit a hierarchical cumulative probit model with question predictability, ellipsis (treatment-coded), and their interaction as population-level effects.[9] We also added crossed group-level intercepts and slopes corresponding to all population-level effects for both items and participants (Barr et al., 2013). This analysis confirmed that elided variants showed a strong question-predictability effect ($\Delta = 2.01, CI(\Delta) = [0.93, 2.96], P(\Delta > 0) = 1$) and that unelided variants were significantly more acceptable than their sluiced counterparts ($\Delta = 1.38, CI(\Delta) = [1.03, 1.74], P(\Delta > 0) = 1$). The interaction between the two, however, was not significant: $\Delta = -0.73, CI(\Delta) = [-2.33, 0.91], P(\Delta < 0) = 0.83$. In other words, the effect of question predictability was not limited to, or significantly more pronounced for, sluicing compared to unelided questions.

[9]This type of model is particularly well-suited for accounting for the possibility of ceiling and floor effects (e.g., Fernandez, Liu, & Costilla, 2019).

6.6.3 Discussion

The results from Experiment 4 confirm that the acceptability of our experimental items is affected by question predictability. They do not, however, rule out the possibility that the effect of question predictability affects unelided variants just as much as it does sentences that involve sluicing. While there is a numerical interaction between ellipsis and question predictability in the right direction (unelided variants were numerically affected less by question predictability), this effect did not reach statistical significance.

One potential concern is that the question predictability measure failed to identify many high-predictability items, leading to a lack of statistical power towards the upper end of the question-predictability scale (see Figure 6.8).[10] This is rooted in the way question predictability was operationalized in Experiment 3: the forced-choice passage-completion paradigm yields a normalized probability distribution in which the scores of all questions in a given context sum to 1. As a result, high scores are only possible when all but one question are considered highly unlikely. It is therefore possible that a different operationalization of question predictability may lead to more statistical power and thus might have a better chance of detecting a sluicing specific effect if one does, in fact, exist. As things stand, however, our results cannot rule out the possibility that the effect of question predictability reflects a general pragmatic effect that is not specific to sluicing.

6.7 General Discussion

6.7.1 Sluicing in the face of extreme lexical mismatches

Drawing on evidence from 4 experiments, this chapter has explored the possibility of sluicing under extreme mismatches between the elided material and its antecedent. One key

[10] Thanks are due to Dan Lassiter for pointing this out to me.

finding from this investigation is that such uses of sluicing can be as highly acceptable despite the lexical mismatches they incur. In this section, I will discuss the implications of this state of affairs for theories of sluicing. Consider the following examples, which were among the most highly rated items across the experiments reported in this chapter:

(185) a. Can I get a few autographs? —Sure, how many (do you want/need)?

b. I think the 49ers need a new quarterback. —Agreed, but who (should it be)?

c. Can I borrow your textbook over the weekend? —Why (do you want to borrow it)?

(186) a. I can't see your parents in the audience. Did you not tell them about your performance today? —I did, but I forgot to tell them where (it would take place).

b. When you call 911 about an emergency, the first thing they ask is where (you are located).

c. Regarding Trump's impeachment, the only question is when (he will be impeached).

The fact that sluicing is felicitous in these contexts is problematic for IDENTITY theories of any kind. First, purely semantic formulations, such as Merchant's (2001) e-GIVENness, classify them as ungrammatical because the elided material and the antecedent do not entail each other. Secondly, purely lexico-syntactic formulations, such as Rudin's (2019) account, falsely classify them as ungrammatical due to the lexical mismatches they incur. Even the "eventive core" restriction, which Rudin proposes in order to capture the possibility of mismatches in modality, polarity, tense, finiteness, and the like, is of little use here. This provision limits the domain of IDENTITY to the highest elided *vP* (which Rudin calls the "eventive core"), but the lexical mismatches in (185) and (186) do fall squarely into that privileged part of the ellipsis site and are thus prohibited under his account. Finally, hybrid IDENTITY proposals that reference both

syntactic and semantic levels of representation fail with respect to these examples because they violate both types of IDENTITY. For example, Chung (2006) adopts both e-GIVENness and the lexico-syntactic condition known as the "No New Words" constraint, both of which are clearly violated by the examples in (185) and (186).

It is important to emphasize at this point that the mismatches we considered in this chapter also violate Chung's (2013) "limited syntactic IDENTITY" account. According to this account, the No New Words constraint only applies to lexical elements that either assign case to the remnant wh-phrase or else determine the argument structure the remnant participates in. This restriction made it possible to permit the mismatches discussed in Chapter 5, which were due to tough movement and the active/passive voice alternation. The mismatches we considered here, however, remain problematic for Chung (2013) for two reasons. First, her account does incorporate e-GIVENness, which is violated independently of the No New Words constraint and the "special heads" restriction. Secondly, while some of the mismatches that proved to be highly acceptable are permitted under Chung's limited syntactic IDENTITY condition, some are clearly not exempt from it. For example, in (185b) the sluicing remnant *who* serves as an argument to the elided predicate and thus requires an identical correlate in the antecedent clause, which it does not have. As a result, the example should be ungrammatical according to Chung (2013), but it was judged to be highly acceptable nonetheless.

Back in Chapter 2 I outlined a tradition of theorizing about ellipsis that revolves around the incremental fine-tuning of the definition of IDENTITY in order to capture increasingly large sets of otherwise problematic observations. As a result, IDENTITY theorists have carved out numerous exceptions to accommodate permissible mismatches, such as ones involving "vehicle change" or syntactic traces (Chung, 2006; Rudin, 2019).[11] The examples in (185) and (186), however, do not lend themselves to this strategy because they reflect a heterogeneous set of mismatches that cannot be captured in a small number of exceptions. This problem is exacerbated by the

[11] See Chapter 2 for arguments that this approach is problematic on theoretical grounds.

fact that our experiments did not just reveal highly acceptable mismatches, but also completely unacceptable ones as well as marginally (un)acceptable cases. Any amendments to the definition of IDENTITY that would capture the acceptable cases would thus likely overgenerate with respect to the marginal and fully unacceptable cases. The next section considers the implications of the gradience in acceptability that our experiments revealed in more detail.

6.7.2 Variable acceptability

Besides the existence of highly acceptable mismatches, as in (185) and (186), another key finding from our investigation is that there is a tremendous amount of variance in acceptability across items. As argued in detail in Chapter 2, this finding is inherently problematic for IDENTITY theories, which aim to classify sentences as either grammatical or ungrammatical. Referential theories of ellipsis, on the other hand, naturally predict gradience since the theoretical constructs they leverage, such as salience and plausibility, are themselves inherently gradient in nature. According to such theories, any deviation from the meaning introduced by the antecedent requires some degree of inference, and the ability to draw the relevant inference—and thus whether or not the use of ellipsis is felicitous in such contexts—depends on factors that are themselves poorly understood to date. The challenge for advocates of referential theories of ellipsis is therefore to identify testable predictions about the acceptability of inferentially resolved ellipsis. The issue at hand can be illustrated with the following minimal pair:

(187) a. Regarding Trump's impeachment, the only question is when (he will be impeached).

b. Regarding Trump's impeachment, the only question is who #(will vote for it).

In Experiment 3 we attempted to explain the gradience in acceptability we had identified in Experiment 2 based on the hypothesis that more predictable questions would be easier to infer.

We operationalized question predictability in terms of a forced-choice passage completion task. This task succeeded in capturing a decent amount of variability, but it was inherently limited in that it failed to identify but a few highly predictable questions in the set of experimental items. As a result, there was limited statistical power towards the upper end of the question-predictability scale and most measures were clustered around the lower end of the scale. Despite this shortcoming, question predictability was significantly correlated with the acceptability of the sluiced variants. This predictability effect did not, however, affect sluiced variants significantly more than their unelided variants, and we thus have no evidence that the effect is ellipsis-specific.

While it is possible that an improved operationalization of the concept of question predictability could reveal a sluicing-specific component (after all, the relevant interaction term was trending in the right direction numerically), it is also possible that question predictability simply does not play a role in the gradience in acceptability associated with nominal-antecedent sluicing. In fact, there are ellipsis-independent reasons to believe that the predictability or salience of the intended discourse referent is not sufficient for ensuring that the use of referring expressions is felicitous. Recall in this context the following example, repeated from (87), which is due to Barbara Partee:

(188) a. I dropped ten marbles and I found all but one of them. It must be under the sofa.

b. I dropped ten marbles and I found only nine of them. # It must be under the sofa.

In (188a), *it* felicitously refers to the tenth marble, which is explicitly introduced into the discourse model by the antecedent NP *one of them*. By contrast, the context in (188b) does not explicitly mention the missing marble and it must therefore be inferred. The fact that this inferential use of the pronoun *it* is infelicitous, however, is surprising given the fact that several aspects of the context conspire to make it maximally salient and predictable: not only does the context plausibly raise the question *Did you find the marbles you dropped?*, the focus construction *only nine of them* specifically shifts attention to the last missing marble. As a result, if inferential reference

resolution were reducible to the accessibility of the intended referent, we should expect the use of *it* to be perfectly felicitous in this context, but it clearly is not. I suspect that the pronoun *it* requires not just a salient referent, but further presupposes that that referent has previously been established as topical, and that explicitly denoting it in a preceding NP helps serve that function.

If sluicing is a form of discourse reference, it is possible that inferential sluicing of the kind we examined in this chapter is likewise not reducible to the accessibility of the intended referent. Fully understanding the gradience associated with inferentially resolved sluiced questions will require a more detailed understanding of the intersection between concepts such as salience, topichood, and predictability.

6.8 Conclusion

Across a series of 4 experiments, we investigated the possibility of inferentially resolved sluicing. We demonstrated that sluicing can be highly acceptable even when the elided material deviates substantially from its antecedent, both semantically and syntactically. We further found variability across items, which raises an important new challenge for all existing theories of sluicing. While the existence of such gradience is inherently consistent with referential theories of sluicing, our investigation highlights the challenge of explaining this variability on an item-by-item basis. IDENTITY theories, on the other hand, are challenged in a more fundamental way by our findings. These findings suggest that the No New Words constraint does not generalize beyond the narrow set of preposition-stranding cases that originally motivated the proposal, and further suggest that semantic IDENTITY conditions like e-GIVENness, which continue to play a central role in most IDENTITY proposals to date, are overly restrictive. Most importantly, the heterogeneous class of mismatches our investigation examined does not lend itself to the fine-tuning approach that is prevalent among IDENTITY theorists.

6.9 Note regarding co-authorship

The material presented in this chapter reflects collaborative work with Andy Kehler and is presently being prepared for publication.

Chapter 7

Conclusion

7.1 Summary

After a brief synopsis of the thesis in Chapter 1, I began in Chapter 2 by providing an overview of the existing literature on ellipsis. I suggested that the theoretical landscape can be productively construed in terms of two theoretical camps: IDENTITY theories, which posit that the use of ellipsis is felicitous only if the elided material corresponds to an identical linguistic antecedent in the context; and referential theories, which reject the notion of a special-purpose IDENTITY condition and instead assume that ellipsis is a form of discourse reference and thus governed by the same underlying mechanisms as pronouns and other discourse-referential devices. I argued that these two positions make fundamentally distinct architectural assumptions and presented arguments for and against both. IDENTITY theories straightforwardly explain the tight correspondence between the ellipsis site and its antecedent that holds in many cases, but all existing definitions of IDENTITY—despite decades-long fine-tuning efforts—continue to be subject to counterexamples in which the elided material and its antecedent are mismatched in various ways. Referential theories, on the other hand, are consistent with the possibility of mismatch, and further provide straightforward explanations for a series of analogies between

ellipsis and other discourse-referential devices, including the possibility of inferential ellipsis resolution. I argued that the central challenge for the referential approach is to develop predictive theories that explain which factors facilitate and constrain inferential ellipsis resolution and under what conditions it is felicitous to elide material that has to be inferred. Against this theoretical backdrop, the subsequent chapters were designed to speak to both types of theories by focusing on a variety of mismatch phenomena that (i) provide new adequacy criteria for the development of IDENTITY theories, and (ii) contribute insights into the inferential mechanisms that are involved in the handling of mismatch cases according to referential theories.

More specifically, Chapters 3 and 4 focused on VP-ellipsis and examined two types of mismatch: voice mismatches in the context of the Recycling Hypothesis, and a novel type of lexical mismatch that arises from the presence of indirect speech acts. Chapters 5 and 6 then pivoted to sluicing and focused on mismatches due to tough movement and the voice alternation, as well as more extreme lexical mismatches in which the elided clause deviates both syntactically and semantically from its antecedent. Each of those investigations revealed highly acceptable instances of ellipsis use despite the presence of extreme mismatch between the elided material and its antecedent, but they also uncovered a tremendous amount of gradience, which underscores the need for explanatory constructs that can capture such gradience in a natural way.

Taken together, these results strengthen the case for referential theories of ellipsis with respect to both VP-ellipsis and sluicing. Not only do they raise a series of novel challenges for existing IDENTITY accounts by showcasing acceptable uses of ellipsis despite extreme mismatch, but the nature of those mismatches suggests that elliptical utterances are interpreted inferentially with the use of world knowledge, which is a core prediction of the referential approach. In the remainder of this chapter, I will briefly outline some ways in which the these findings have theoretical implications beyond the study of ellipsis and touch on some salient issues in the literature that have not been discussed in previous chapters.

7.2 Implications beyond the theory of ellipsis

Because the ellipsis literature is intimately intertwined with many other literatures in linguistics, evidence for or against certain theories of ellipsis can have implications far beyond the study of ellipsis *per se*. This is particularly clear with respect to theories of phenomena that draw on evidence from ellipsis, since the epistemic force of those arguments often relies on an IDENTITY-based approach to ellipsis. While I leave a more comprehensive review of linguistic research that relies on particular theoretical approaches to ellipsis to future work, I will provide two examples to illustrate the general idea.

To appreciate the first example, recall from Chapter 2 that Merchant (2013a) lays out the implications of adopting a syntactic IDENTITY theory of ellipsis for the theory of polarity items and the syntactic representation of the negative determiner *no* (see Section 2.1.2 for details).

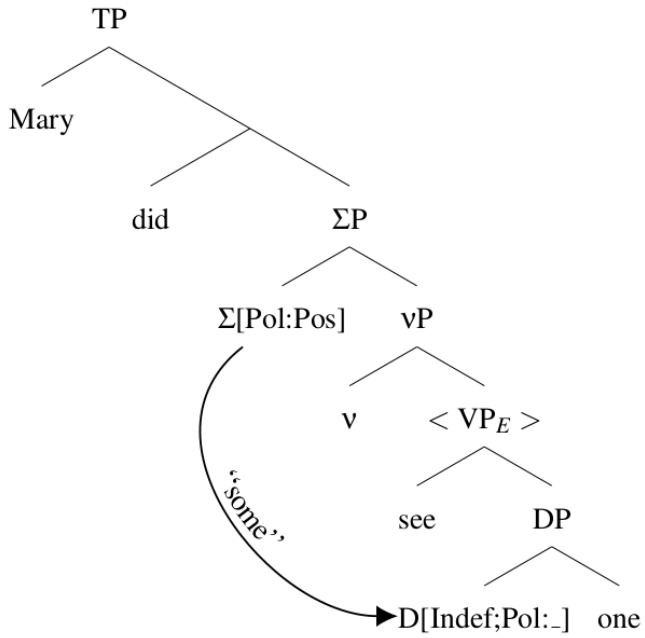
(189) a. John doesn't see anyone, but Bill does (see someone). (Sag, 1976, ex. 2.3.39)

b. John saw someone, but Mary didn't (see anyone). (Merchant, 2013a, ex. 2)

c. We haven't decided to blacklist any firms. But there's a chance we might (blacklist some firms). (Hardt, 1993, ex. 68)

d. I could find no solution, but Holly might (find a solution). (Johnson, 2001, ex. 107)

As I detailed in Section 2.1.2, Merchant (2013a) argues that holding on to a lexico-syntactic IDENTITY condition for ellipsis in light of the observations in (189) requires adopting a view of polarity items that construes the mismatching elements as underlyingly identical. Specifically, he proposes that polarity items are abstract entities at the level of syntax and receive their mismatching surface forms in agreement with a polarity phrase—Σ*P*—that is itself located outside the ellipsis clause and therefore exempt from the IDENTITY condition. This analysis is illustrated in the following syntactic tree, repeated from (35).

(190) 

(see Merchant, 2013a, ex. 3)

The choice of adopting a syntactic IDENTITY theory of ellipsis thus may have far-reaching implications beyond the theory of ellipsis, not only with respect to the nature of polarity items, but also the interface between syntactic and morphological levels of representation, the nature of syntactic agreement, and so forth. Following the same logic, then, evidence *against* a lexico-syntactic IDENTITY requirement, like the experimental results reported throughout this thesis, serve as an argument against the need to adopt "distributed" syntactic representations of polarity items. The same reasoning applies to the syntactic representation of de-verbal NPs, which, as I discussed in Section 2.1.2, have been argued to underlyingly contain the VPs they derive from in order to explain the possibility of nominal-antecedent VP-ellipsis (Fu et al., 2001; Johnson, 2001; Merchant, 2013a): if ellipsis is *not* subject to an IDENTITY condition, there is no reason to stipulate silent VP representations, especially in the absence of independent evidence pointing to that conclusion.

A second example comes from Hartman (2011), who draws conclusions about the nature of syntactic traces and the syntax-semantics interface based primarily on evidence from ellipsis. Specifically, he argues against the proposal that head movement is purely phonological in nature

and does not leave behind any syntactic traces that affect interpretation. While the details of Hartman's proposal would take us too far afield, it is important to emphasize that the validity of his argument depends on the assumption that ellipsis is governed by an IDENTITY condition (along with several other assumptions about the nature of ellipsis that need not concern us here).[1] In his own words:

> Using the identity condition on ellipsis as a diagnostic tool, I show that A-traces, $\bar{A}$-traces, and traces of head movement are all interpreted as bound variables.
>
> (Hartman, 2011, p. 367)

As in the case of Merchant's analysis of polarity items, Hartman's reliance on an IDENTITY approach to ellipsis makes his proposal vulnerable to challenges from the "outside:" evidence that calls into question whether ellipsis is, in fact, governed by an IDENTITY condition, will likewise threaten to undermine his arguments about the nature of movement, syntactic traces, and the syntax-semantics interface.

7.3 Reducing ellipsis to analogical reasoning?

Before closing, I would like to briefly revisit an approach to ellipsis that is most commonly attributed to Culicover and Jackendoff (2005, 2012) and that I have thus far only mentioned in passing (see also Goldberg & Perek, 2019; Kim & Nykiel, 2020). This approach is sometimes referred to as the "interpretive approach" and has often been grouped—I think mistakenly—together with referential approaches to ellipsis (e.g., Lipták, 2015; Wood, Barros, & Sigurdsson, 2020). It is characterized by two core assumptions. First, it assumes that the ellipsis site does not contain any syntactic structure—not even a null pro-form. Sluiced questions, for example, are analyzed as consisting of nothing more than the remnant wh-phrase, which is broadly consistent

[1] See Schoorlemmer and Temmerman (2012) for an IDENTITY-based argument that supports the opposite conclusion, i.e. that head movement *is* primarily phonological in nature.

with Culicover and Jackendoff's (2005) *Simpler Syntax* framework. Secondly, elliptical utterances are assumed to acquire their meaning by recruiting domain-general reasoning mechanisms that are independently needed for the perception of similarities and differences between objects during analogical reasoning (e.g., Sagi, Gentner, & Lovett, 2012), as well as for the meaning of expressions like *same*, *except*, anaphoric *one* (see also Goldberg & Michaelis, 2017), *vice versa*, and so forth. In that sense, there is nothing "special" about ellipsis: its interpretation involves the recognition of the same semantic relation, which Culicover and Jackendoff term "SAME-EXCEPT," as utterances without ellipsis and even entirely non-linguistic cognitive processes.

This position has several aspects in common with referential theories of ellipsis. Both allow for the possibility that domain-general reasoning mechanisms are recruited for the interpretation of ellipsis, and both reject the notion that elliptical utterances contain fully formed syntactic structure that is silent but otherwise identical to their unelided counterparts. They also both reject the assumption that ellipsis requires a linguistic antecedent that is identical to the elided material. However, there is a fundamental difference between the two as well, and it is precisely that point of divergence that creates what I think are fatal problems for Culicover and Jackendoff's approach.

Recall from Chapter 2 that important evidence that elliptical utterances contain silent pro-forms and engage the referential system comes from the fact that they exhibit a series of diagnostic properties associated with discourse reference: the possibility of exophora, multiple "split" antecedents, non-local antecedents, cataphora, the ability to trigger "sloppy" interpretations, and the possibility of inferential ellipsis resolution. Since Culicover and Jackendoff's account does not posit the presence of pro-forms or that elliptical utterances engage the referential system in any way, it does not explain why ellipsis exhibits these diagnostic properties.

Furthermore, because it reduces ellipsis interpretation to a domain-general mechanism that handles both elliptical and non-elliptical utterances and provides a unified treatment of several different types of ellipsis (including Bare Argument Ellipsis, sluicing, VP-ellipsis, and Gapping),

it fails to capture important disanalogies within this class. First, recall that unelided variants do not exhibit the ability to trigger sloppy interpretations in the same way that elliptical utterances do, as shown in (191), repeated from (69).[2]

(191) a. When Harry drinks, I always conceal my belief that he shouldn't (drink). But when he gambles, I often can't (conceal my belief that he shouldn't gamble).

b. When Harry drinks, I always conceal my belief that he shouldn't drink. But when he gambles, I often can't #(conceal my belief that he shouldn't gamble).

When the verb *drink* is elided, the subsequent ellipsis site can receive a sloppy interpretation (where *drink* is replaced by *gamble*), but the same reading is unavailable (or at least less available) when the initial ellipsis site is replaced with an overt instance of the verb *drink*. According to Culicover and Jackendoff's theory of ellipsis, however, both of those utterances are interpreted through the same analogical reasoning mechanism, and as a result the contrast between them remains unexplained. Furthermore, since elliptical utterances do not engage the referential system according to that account, it cannot explain the fact that non-elliptical referring expressions, such as *itself* in (192), exhibit the same behavior, or why—once again—replacing the referring expression with a synonymous but non-referential form causes the sloppy reading to disappear, as shown in (192b):

(192) a. 5 is equal to itself and 7 is (equal to itself, i.e. 7), too.

(adapted from Rooth, 1992, ex. 5)

b. 5 is equal to 5 and 7 is #(equal to itself, i.e. 7), too

A second, perhaps even more severe, problem arises from the fact that Culicover and Jackendoff (2012) provide a unified treatment for various types of ellipsis, including VP-ellipsis,

[2]This type of example was first discussed by Hardt (1994) and later re-discovered by Schwarz (2000).

sluicing, and Gapping. If they all are governed by the same underlying mechanism, we would expect them to pattern together with respect to diagnostic properties of discourse reference. Recall from Section 2.2.2, however, that while VP-ellipsis and sluicing exhibit all of them, Gapping does not appear to exhibit any of them:

(193) a. Exophora: [Context: right after a car runs a red light.] One pedestrian to another: Yesterday, a TRUCK #(ran) a red light.

b. Split antecedents: Leslie saw the first car coming and Beto heard it. Neither of them #(saw/heard, respectively) the second one.

c. Non-local antecedents: Nina called her father on Monday. On Tuesday she was busy all day. That's why her sister #(called) her mother.

d. Cataphora:

(i) Even though Susan #(accepted) Bob's apology, Jessie didn't accept Bill's.

(ii) Susan #(accepted) Bob's apology, and Jessie accepted Bill's.

e. Triggering sloppy interpretations: The women all called their friends and the men texted theirs. Specifically, Susan said that she called her friends and her friends (called) theirs, and Jack did #(say that he texted his friends and his friends texted theirs), too.

f. Inferred referents: Irv and Mary want to dance together, and Jack #(wants to dance with) Sue.

It therefore seems to me that Culicover and Jackendoff's attempt to provide a unified account of different types of ellipsis (as well as linguistic behavior beyond ellipsis and indeed even non-linguistic aspects of cognition) does more harm than good: while it captures some commonalities across these empirical domains, there are also important disanalogies that suggest to me that they are unlikely to be governed by the same underlying mechanism. Furthermore, their treatment of ellipsis obscures the fact that VP-ellipsis and sluicing exhibit hallmark features of discourse

reference, which makes it all the more surprising that Culicover and Jackendoff's account and referential accounts tend to be grouped together in the literature.

7.4 Final thoughts

Ellipsis is an intriguing phenomenon in its own right. It is pervasive across the world's languages and its distribution is so complex that after more than five decades and countless theses and articles, we still do not have an adequate description of its distribution, let alone an explanation for it. Furthermore, the study of ellipsis has a way of challenging the way we analyze other linguistic phenomena, including the syntactic structure of sentences and issues at the syntax-semantics interface (Johnson, 2008). We have seen examples of this at various points in this thesis, spanning phenomena like Case connectivity (Section 2.3.2), argument-structure alternations (Chapters 3 and 5), and the role of inference in language comprehension (Chapters 4 and 6).

To me personally, however, those intriguing aspects pale in comparison to the insights ellipsis promises to deliver more generally about the way languages rely on information that is made available by the context of utterance. Kehler and Rohde (2013) make a similar point about the study of pronouns, which they consider "fruit flies" in the study of discourse processing mechanisms. In the same way, I view ellipsis not just as an intriguing puzzle in its own right, but as a window into the mechanisms that support the use of shared contextual information, which is a fundamental aspect of human language. To illustrate this point, suppose that I were to quiz you, the reader, about this thesis after you have finished reading it and that we both tried our best to avoid using any context-dependent expressions:

(194) Me: What do you remember about this thesis?
You: Well, this thesis is about ellipsis.
Me: What types of ellipsis is this thesis about?

You: This thesis is about VP-ellipsis and sluicing.

Me: Does this thesis report any experiments?

You: This thesis does report experiments.

Me: How many experiments does this thesis report?

You: This thesis reports 10 experiments.

Clearly, this is an extremely unnatural conversation. It is highly redundant in a variety of ways because we refuse to take advantage of the common ground of information we are both mutually aware of, which expands as the dialogue unfolds. Unsurprisingly, then, the same exchange becomes much more natural once the redundant information is replaced by pronouns and the use of ellipsis wherever possible:

(195) Me: What do you remember about this thesis?

You: Well, it is about ellipsis.

Me: What types of ellipsis?

You: VP-ellipsis and sluicing.

Me: Does it report any experiments?

You: It does.

Me: How many?

You: 10.

The difference between (194) and (195) is striking because it touches upon a core property of human languages: they are designed to take advantage of the shared information between speakers and their addressees. Not only does this make linguistic communication efficient, it also allows interlocutors to coordinate their beliefs and intentions and to jointly control the topic and flow of the discourse. That to me is the single most fascinating aspect about ellipsis: it provides a window into the way languages make use of information that is available in the context of

utterance, and how the language-context interface interacts with grammatical constraints on the use of context-dependent expressions.

Bibliography

Anand, P., & Hardt, D. (2016). Antecedent selection for sluicing: Structure and content. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 1234–1243).

Anand, P., & McCloskey, J. (2015). Annotating the implicit content of sluices. In *Proceedings of the 9th Linguistic Annotation Workshop* (pp. 178–187).

AnderBois, S. (2010). Sluicing as anaphora to issues. In *Proceedings of Semantics and Linguistic Theory* (Vol. 20, pp. 451–470).

AnderBois, S. (2014). The semantics of sluicing: Beyond truth conditions. *Language*, *90*(4), 887–926.

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, *24*(1), 65–87.

Ariel, M. (1991). The function of accessibility in a theory of grammar. *Journal of Pragmatics*, *16*(5), 443–463.

Arregui, A., Clifton, C., Frazier, L., & Moulton, K. (2006). Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language*, *55*(2), 232–246.

Baker, A. C. (2007). Discourse Coherence and VP Ellipsis With Split Antecedents. In *Proceedings of the Sixteenth Amsterdam Colloquium*.

Barker, C. (2013). Scopability and sluicing. *Linguistics and Philosophy*, *36*(3), 187–223.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Barros, M. (2014). *Sluicing and identity in ellipsis* (Doctoral dissertation, Rutgers University).

Beecher, H. (2007). Pragmatic Inference in the Interpretation of Sluiced Prepositional Phrases. In *Proceedings of the Fifth University of Cambridge Postgraduate Conference in Language Research* (pp. 9–16). Citeseer.

Bos, J. (1993). VP Ellipsis in a DRT-implementation. In *Sixth Conference of the European Chapter of the Association for Computational Linguistics*, Utrecht, The Netherlands: Association for Computational Linguistics. Retrieved from https://www.aclweb.org/anthology/E93-1050

Brown-Schmidt, S., Byron, D. K., & Tanenhaus, M. K. (2005). Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language*, *53*(2), 292–313.

Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28.

Bürkner, P.-C. (2018). Advanced bayesian multilevel modeling with the r package brms. *The R Journal*, 1–15. Retrieved from https://arxiv.org/abs/1705.11123.

Chao, W. (1988). *On ellipsis* (Doctoral dissertation, University of Massachusetts Amherst). Retrieved from https://scholarworks.umass.edu/dissertations/AAI8710433

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.

Chung, S. (2006). Sluicing and the lexicon: The point of no return. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* (Vol. 31, pp. 73–91). Berkeley, CA.

Chung, S. (2013). Syntactic identity in sluicing: How much and why. *Linguistic Inquiry*, *44*(1), 1–44.

Chung, S., Ladusaw, W. A., & McCloskey, J. (1995). Sluicing and logical form. *Natural Language Semantics*, *3*(3), 239–282.

Chung, S., Ladusaw, W., & McCloskey, J. (2011). Sluicing (:) between structure and inference. *Representing language: Essays in honor of Judith Aissen*, 31–50.

Clark, H. H. (1975). Bridging. In B. L. Nash-Webber & R. C. Schank (Eds.), *Theoretical Issues in Natural Language Processing*.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding*.

Culicover, P. W., & Jackendoff, R. (2005). *Simpler Syntax*. Oxford University Press on Demand.

Culicover, P. W., & Jackendoff, R. (2012). Same-except: A domain-general cognitive relation and how language expresses it. *Language*, *88*(2), 305–340.

Dalrymple, M., Shieber, S. M., & Pereira, F. C. (1991). Ellipsis and higher-order unification. *Linguistics and Philosophy*, *14*(4), 399–452.

Elbourne, P. (2001). E-type anaphora as NP-deletion. *Natural Language Semantics*, *9*(3), 241–288.

Elbourne, P. (2008). Ellipsis sites as definite descriptions. *Linguistic Inquiry*, *39*(2), 191–220.

Fernandez, D., Liu, I., & Costilla, R. (2019). A method for ordinal outcomes: The ordered stereotype model. *International Journal of Methods in Psychiatric Research*, *28*(4).

Fiengo, R., & May, R. (1994). *Indices and identity*. MIT press.

Fox, D. (1999). Focus, parallelism and accommodation. In *Semantics and linguistic theory* (Vol. 9, pp. 70–90).

Frazier, L. (2013). A Recycling Approach to Processing Ellipsis. In L. L.-S. Cheng & N. Corver (Eds.), *Diagnosing Syntax* (pp. 485–501).

Frazier, L., & Duff, J. (2019). Repair or accommodation? Split antecedent ellipsis and the limits of repair. *Glossa: a Journal of General Linguistics*, *4*(1).

Fu, J., Roeper, T., & Borer, H. (2001). The VP within Process Nominals: Evidence from Adverbs and the VP Anaphor Do-So. *Natural Language & Linguistic Theory*, *19*(3), 549–582.

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science*, *44*(3), e12814.

Futrell, R., & Levy, R. (2017). Noisy-context surprisal as a human sentence processing cost model. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (pp. 688–698).

Geach, P. T. (1962). Reference and generality: An examination of some medieval and modern theories.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189–211.

Gibson, E., Futrell, R., Piandadosi, S. T., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*. doi:https://doi.org/10.1016/j.tics.2019.02.003

Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, *14*(3), 225–248.

Ginzburg, J. (1992). *Questions, queries and facts: A semantics and pragmatics for interrogatives* (Doctoral dissertation, Stanford University).

Ginzburg, J. (2012). *The interactive stance*. Oxford University Press.

Ginzburg, J., & Miller, P. (2019). Ellipsis in head-driven phrase structure grammar. In J. van Craenenbroeck & T. Temmerman (Eds.), *Handbook of ellipsis*. Oxford: Oxford University Press.

Ginzburg, J., & Sag, I. A. (2000). *Interrogative investigations*. Stanford: CSLI publications.

Givón, T. (1990). *Syntax: A functional typological introduction*. Amsterdam: John Benjamins.

Goldberg, A. E., & Michaelis, L. A. (2017). One Among Many: Anaphoric One and Its Relationship With Numeral One. *Cognitive Science*, *41*, 233–258.

Goldberg, A. E., & Perek, F. (2019). Ellipsis in head-driven phrase structure grammar. In J. van Craenenbroeck & T. Temmerman (Eds.), *The Oxford Handbook of Ellipsis*. doi:https://doi.org/10.1093/oxfordhb/9780198712398.013

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, *17*(3), 311–347.

Grant, M., Clifton Jr, C., & Frazier, L. (2012). The Role of Non-Actuality Implicatures in Processing Elided Constituents. *Journal of Memory and Language*, *66*(1), 326–343.

Grosz, B. J. (1977). *The Representation and Use of Focus in Dialogue Understanding*. SRI International Menlo Park CA, United States.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, *21*(2), 203–225.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 274–307.

Hale, K., & Keyser, S. J. (1993). On Argument Structure and the Lexical Expression of Syntactic Relations. In K. Hale (Ed.), *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger* (pp. 53–109). MIT Press.

Hankamer, J. (1978). On the Nontransformational Derivation of Some Null VP Anaphors. *Linguistic Inquiry*, *9*(1), 66–74.

Hankamer, J., & Sag, I. A. (1976). Deep and surface anaphora. *Linguistic inquiry*, *7*(3), 391–428.

Hardt, D. (1990). A corpus-based survey of VP ellipsis. *Ms., University of Pennsylvania*.

Hardt, D. (1992). VP ellipsis and Semantic Identity. In *Semantics and Linguistic Theory* (Vol. 2, pp. 145–162).

Hardt, D. (1993). *Verb phrase ellipsis: Form, meaning, and processing* (Doctoral dissertation, University of Pennsylvania).

Hardt, D. (1994). Sense and Reference in Dynamic Semantics. In *Proceedings of the Ninth Amsterdam Colloquium* (pp. 333–348).

Hardt, D. (1999). Dynamic interpretation of verb phrase ellipsis. *Linguistics and philosophy*, 185–219.

Hardt, D., Anand, P., & McCloskey, J. (2020). *Sluicing: Antecedence and the landscape of mismatch*. Talk presented at Experimental and Corpus-based Approaches to Ellipsis. Retrieved from http://drehu.linguist.univ-paris-diderot.fr/ecbae-2020/fichiers/abstracts/hardt.pdf

Hartman, J. (2011). The semantic uniformity of traces: Evidence from ellipsis parallelism. *Linguistic Inquiry*, *42*(3), 367–388. doi:10.1162/LING_a_00050

Hartman, J. (2009). When e-givenness over-predicts identity. In *Fourth Brussels Conference on Generative Linguistics Ellipsis Workshop* (Vol. 4).

Hinds, J., & Okada, N. (1975). Backward pronominalization across coordinate structures. *Linguistic Inquiry*, 330–335.

Jäger, G. (2001). Indefinites and sluicing. A type logical approach. In *Proceedings of the 13th Amsterdam Colloquium*.

Jäger, G. (2005). *Anaphora and type logical grammar*. Springer Science & Business Media.

Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*, *10*(4), 409–416.

Johnson, K. (2001). What vp ellipsis can do, and what it can't, but not why. In *The Handbook of Contemporary Syntactic Theory* (pp. 439–479). Wiley Online Library.

Johnson, K. (2008). Introduction. In K. Johnson (Ed.), *Topics in Ellipsis* (Chap. 1, pp. 1–14). Cambridge University Press.

Karttunen, L. (1969). Pronouns and Variables. In *Fifth Regional Meeting of the Chicago Linguistics Society* (pp. 108–115).

Karttunen, L. (1976). Discourse referents. In *Notes from the linguistic underground* (pp. 363–385). Brill.

Katzir, R. (2007). Structurally-defined alternatives. *Linguistics and Philosophy*, *30*(6), 669–690.

Kehler, A. (1993a). A discourse copying algorithm for ellipsis and anaphora resolution. In *Proceedings of the sixth conference on european chapter of the association for computational linguistics* (pp. 203–212). Association for Computational Linguistics.

Kehler, A. (1993b). The effect of establishing coherence in ellipsis and anaphora resolution. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics* (pp. 62–69).

Kehler, A. (2000). Coherence and the resolution of ellipsis. *Linguistics and Philosophy*, *23*(6), 533–575.

Kehler, A. (2002a). Another problem for syntactic (and semantic) theories of vp-ellipsis. *Snippets*, *5*, 11–12.

Kehler, A. (2002b). *Coherence, Reference, and the Theory of Grammar*. CSLI Publications Stanford, CA.

Kehler, A. (2016). On qud-based licensing of strict and sloppy ambiguities. In *Semantics and linguistic theory* (Vol. 25, pp. 512–532).

Kehler, A. (2017). *Opportunities and challenges for a referential theory of VP-ellipsis*. Talk presented at the Workshop on Experimental and Corpus-based Approaches to Ellipsis, LSA Institute, Lexington, KY.

Kehler, A. (2019). Ellipsis and discourse. In J. van Craenenbroeck & T. Temmerman (Eds.), *The Oxford Handbook of Ellipsis* (pp. 314–341). Oxford: Oxford University Press.

Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, *39*(1-2), 1–37.

Kehler, A., & Ward, G. (2007). Event reference and semantic transparency. In *Western conference on linguistics* (Vol. 18, pp. 115–127).

Kertz, L. (2008). Focus Structure and Acceptability in Verb Phrase Ellipsis. In N. Abner & J. Bishop (Eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics* (pp. 283–291).

Kertz, L. (2010). *Ellipsis reconsidered* (Doctoral dissertation, University of California, San Diego, San Diego, CA).

Kertz, L. (2013). Verb phrase ellipsis: The view from information structure. *Language*, *89*(3), 390–428.

Khullar, P., Majmundar, K., & Shrivastava, M. (2020). NoEl: An Annotated Corpus for Noun Ellipsis in English. In *Proceedings of The 12th Language Resources and Evaluation Conference* (pp. 34–43).

Kim, C. S., Kobele, G. M., Runner, J. T., & Hale, J. T. (2011). The Acceptability Cline in VP Ellipsis. *Syntax*, *14*(4), 318–354. doi:10.1111/j.1467-9612.2011.00160.x

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, J.-B., & Nykiel, J. (2020). The syntax and semantics of elliptical constructions: A direct interpretation perspective. *Linguistic Research*, *39*.

Klein, E. (1987). VP Ellipsis in DR Theory. In *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers* (8, pp. 161–187).

Kroll, M. (2019). Polarity reversals under sluicing. *Semantics and Pragmatics*, *12*(18), 1–49. doi:http://dx.doi.org/10.3765/sp.12.18

Lasnik, H. (1995). Case and expletives revisited: On greed and other human failings. *Linguistic Inquiry*, 615–633.

Levin, L. (1982). *Sluicing: A lexical interpretation procedure*. MIT Press.

Lewis, D. (1969). *Convention: A philosophical study*. Harvard Univeristy Press.

Lewis, D. (1979). Scorekeeping in a language game. In *Semantics from different points of view* (pp. 172–187). Springer.

Lipták, A. (2015). Identity in ellipsis: An introduction. *Lingua*, (166), 155–171. doi:10.1016/j.lingua.2015.08.003

Mehler, J. (1963). Some Effects of Grammatical Transformations on the Recall of English Sentences. *Journal of Verbal Learning and Verbal Behavior*, *2*(4), 346–351.

Merchant, J. (1999). E-type A'-traces under Sluicing. In *The Proceedings of the Seventeenth West Coast Conference on Formal Linguistics* (Vol. 30, pp. 478–492).

Merchant, J. (2001). *The Syntax of Silence: Sluicing, islands, and the Theory of Ellipsis*. Oxford University Press on Demand.

Merchant, J. (2004). A deletion solution to the sloppy ellipsis puzzle. Talk presented at the 75th Annual Meeting of the Linguistics Society of America. Retrieved from http://home.uchicago.edu/merchant/pubs/LSA2004.handout.pdf

Merchant, J. (2005). *Revisiting syntactic identity conditions*. Talk presented at a Workshop on Identity in Ellipsis, UC Berkeley.

Merchant, J. (2008). An Asymmetry in Voice Mismatches in VP-Ellipsis and Pseudogapping. *Linguistic Inquiry*, *39*(1), 169–179.

Merchant, J. (2013a). Polarity items under ellipsis. In L. L.-S. Cheng & N. Corver (Eds.), *Diagnosing syntax*. Oxford University Press Oxford.

Merchant, J. (2013b). Voice and ellipsis. *Linguistic Inquiry*, *44*(1), 77–108.

Merchant, J. (2019). Ellipsis: A survey of analytical approaches. In J. van Craenenbroek & T. Temmerman (Eds.), *The Oxford Handbook of Ellipsis* (Chap. 1, pp. 19–45). Oxford, United Kingdom: Oxford University Press.

Messick, T. (2012). Ellipsis and reconstruction in tough-infinitives. In *Proceedings of Generative Linguistics in the Old World in Asia IX* (pp. 173–185).

Messick, T., Saab, A., & Vicente, L. (2016). *Deep properties of a surface anaphor: On the contextual restriction of sluicing sites*.

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users.

Miller, P. (2011). The choice between verbal anaphors in discourse. In *Discourse anaphora and anaphor resolution colloquium* (pp. 82–95). Springer.

Miller, P., & Hemforth, B. (2014). *Verb phrase ellipsis with nominal antecedents*.

Miller, P., & Pullum, G. K. (2013). Exophoric VP Ellipsis. In P. Hofmeister & E. Norcliffe (Eds.), *The Core and the Periphery: Data-driven Perspectives on Syntax Inspired by Ivan A. Sag* (pp. 5–32).

Nash-Webber, B. L. (1977). Inference in an Approach to Discourse Anaphora. *Center for the Study of Reading*. Technical Report No. 77.

Nykiel, J. (2010). Whatever Happened to English Sluicing. In *Studies in the History of the English Language V: Variation and Change in English Grammar and Lexicon: Contemporary Approaches* (Chap. 3, Vol. 5, pp. 37–66). Walter de Gruyter.

Paape, D. L. J. F. (2016). Filling the silence: Reactivation, not reconstruction. *Frontiers in Psychology*, *7*, 27. doi:10.3389/fpsyg.2016.00027

Parker, D. (2018). A memory-based explanation of antecedent-ellipsis mismatches: New insights from computational modeling. *Glossa: a journal of general linguistics*, *3*(1).

Parker, D. (2017). *Navigating ellipsis structures in memory: New insights from computational modeling*. LSA workshop: Corpus-based and Experimental Approaches to Ellipsis.

Poppels, T., & Kehler, A. (2018). Overcoming the identity crisis: Novel evidence for a referential theory of verb phrase ellipsis. In *Proceedings of the Annual Meeting of the Chicago Linguistics Society* (Vol. 53, pp. 403–4017).

Poppels, T., & Kehler, A. (2019). Reconsidering asymmetries in voice-mismatched verb phrase ellipsis. *Glossa: A Journal of General Linguistics*, *4*, 1–22. doi:http://doi.org/10.5334/gjgl.738

Postal, P. (1969). Anaphoric islands. In *Proceedings of the Annual Meeting of the Chicago Linguistic Society* (Vol. 5, pp. 205–239).

Prince, E. F. (1981). Towards a taxonomy of given-new information. *Radical pragmatics*.

Prince, E. F. (1986). On the syntactic marking of presupposed open propositions. In *Papers from the parasession on pragmatics and grammatical theory at the 22nd regional meeting of the Chicago Linguistic Society* (pp. 208–222). Chicago, IL: Chicago Linguistics Society.

R Development Core Team. (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.R-project.org

Roberts, C. (1998). *Information structure in discourse: Towards an integrated formal theory of pragmatics*.

Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, *5*(6), 1–69. doi:10.3765/sp.5.6

Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience*, *29*(8), 912–927.

Rønning, O., Hardt, D., & Søgaard, A. (2018). Sluice resolution without hand-crafted features over brittle syntax trees. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)* (pp. 236–241).

Rooth, M. (1992). Ellipsis redundancy and reduction redundancy. In *Proceedings of the stuttgart ellipsis workshop* (Vol. 29). Citeseer.

Ross, J. (1969). Guess who? In *Papers from the fifth regional meeting of the Chicago Linguistic Society.* (pp. 252–286).

Rudin, D. (2019). Head-based syntactic identity in sluicing. *Linguistic Inquiry*, *50*(2), 253–283.

Sag, I. A. (1976). *Deletion and logical form* (Doctoral dissertation, Massachusetts Institute of Technology).

Sag, I. A., & Hankamer, J. (1984). Toward a theory of anaphoric processing. *Linguistics and philosophy*, 325–345.

Sagi, E., Gentner, D., & Lovett, A. (2012). What difference reveals about similarity. *Cognitive science*, *36*(6), 1019–1050.

SanPietro, S., Merchant, J., & Xiang, M. (2012). Accounting for voice mismatch in ellipsis. In *Proceedings of the 30th west coast conference on formal linguistics, ed. by nathan arnett and ryan bennett* (pp. 303–312).

Schachter, P. (1977). Does she or doesn't she? *Linguistic Inquiry*, *8*(4), 763–767.

Schoorlemmer, E., & Temmerman, T. (2012). Head Movement as a PF-Phenomenon: Evidence from Identity under Ellipsis. In *Proceedings of the West Coast Conference on Formal Linguistics. Cascadilla, Somerville, MA*.

Schwarz, B. (2000). *Topics in Ellipsis* (Doctoral dissertation, University of Massachusetts Amherst).

Schwarzschild, R. (1999). GIVENness, AvoidF and Other Constraints on the Placement of Accent. *Natural Language Semantics*, *7*(2), 141–177.

Shibatani, M. (1985). Passives and related constructions: A prototype analysis. *Language*, *61*, 821–848.

Sproat, R., & Ward, G. (1987). Pragmatic considerations in anaphoric island phenomena. In *Chicago linguistic society* (Vol. 23, pp. 321–335).

Steedman, M. J. (1990). Gapping as constituent coordination. *Linguistics and Philosophy*, *13*(2), 207–263.

Tanaka, H. (2011a). Syntactic identity and ellipsis. *The Linguistic Review*, *28*(1), 79–110.

Tanaka, H. (2011b). Voice mismatch and syntactic identity. *Linguistic Inquiry*, *42*(3), 470–490.

Tancredi, C. D. (1992). *Deletion, deaccenting, and presupposition* (Doctoral dissertation, Massachusetts Institute of Technology).

Thoms, G. (2015). Syntactic identity, parallelism and accommodated antecedents. *Lingua*, *166*, 172–198.

Tic Douloureux, P. R. N. (1971). A note on one's privates. In *Studies Out in Left Field: Defamatory Essays Presented to James D. McCawley on the Occasion of His 33rd or 34th Birthday* (pp. 45–52). Edmonton and Champaign: Linguistic Research, Inc.

Tomioka, S. (2008). A step-by-step guide to ellipsis resolution. In K. Johnson (Ed.), *Topics in Ellipsis* (Chap. 9, pp. 210–228). Cambridge University Press Cambridge.

Van Craenenbroeck, J., & Merchant, J. (2013). Ellipsis phenomena. In M. d. Dikken (Ed.), *Cambridge Handbook of Generative Syntax* (Chap. 19, pp. 701–745). Cambridge University Press.

Vicente, L. (2019). Sluicing and its subtypes. In J. van Craenenbroek & T. Temmerman (Eds.), *The Oxford Handbook of Ellipsis* (Chap. 20). Oxford Handbooks. Oxford, United Kingdom: Oxford University Press.

Ward, G., & Kehler, A. (2005). Syntactic form and discourse accessibility. *Anaphora processing: Linguistic, cognitive and computational modelling*, 365–384.

Ward, G., Sproat, R., & McKoon, G. (1991). A pragmatic analysis of so-called anaphoric islands. *Language*, *67*(3), 439–474.

Wasow, T. (1972). *Anaphoric relations in English* (Doctoral dissertation, Massachusetts Institute of Technology).

Webber, B. L. (1978). *A Formal Approach to Discourse Anaphora*. Bolt Beranek and Newman Inc. Cambirdge, MA.

Williams, E. S. (1977). Discourse and logical form. *Linguistic inquiry*, 101–139.

Wilson, D., & Sperber, D. (1979). Ordered entailments: An alternative to presuppositional theories. In C. K. Oh & D. Dinneen (Eds.), *Syntax and semantics XI: Presupposition* (pp. 299–323). Academic Press.

Wood, J., Barros, M., & Sigurdsson, E. F. (2020). Case mismatching in icelandic clausal ellipsis. *Journal of Linguistics*, *56*(2), 399–439.

Xiang, M., & Klafka, J. (2018). *Memory retrieval in comprehension is sensitive to production alternatives*. Poster presented at the 31st Annual CUNY Sentence Processing Conference, Davis, CA.