Poppels, Till & Kehler, Andrew. 2025. Novel cases of sluicing with mismatched antecedents: Theoretical consequences. *Glossa: a journal of general linguistics* 10(1). pp. 1–37. DOI: https://doi.org/10.16995/glossa.15307

Novel cases of sluicing with mismatched antecedents: Theoretical consequences

Till Poppels, Department of Linguistics, UC San Diego, US, tillpoppels@gmail.com

Andrew Kehler, Department of Linguistics, UC San Diego, US, akehler@ucsd.edu

Sluicing is widely regarded to disallow syntactic mismatches at the constructional level, an assumption that has strongly shaped the character of contemporary theories. Here we expand the empirical base for theorizing by examining the felicity of sluicing under two types of constructional mismatches experimentally: tough mismatches and voice mismatches. The results reveal a set of highly felicitous sluices with both types of mismatch, as well as a high degree of variability across individual items. We compare these results against the predictions of a range of analyses and explain their ramifications for the existing theoretical landscape.

1 Introduction

When context allows, natural languages permit speakers to omit linguistic material that expresses information that they nonetheless intend to convey. This design feature allows speakers to minimize articulatory effort, at least in those situations in which the hearer is able to perform the intended meaning recovery. An important class of phenomena that speakers can employ falls under the cover term of ELLIPSIS. Unlike some other forms of content enrichment (implicatures, etc), ellipsis phenomena are among those in which the grammar itself provides a tip-off to the hearer that normally-required linguistic material is in fact missing. Successful interpretation in such cases is enabled by the fact that context allows for the recovery of already activated (precomputed or otherwise predictable, and hence redundant) meanings.

There are many types of ellipsis found in the world's languages, with an impressively diverse set of constraints on their use. A particularly well-studied type is SLUICING (Ross 1969), illustrated in (1).

(1) I remember the paper was about ellipsis, but I don't remember which type of ellipsis (the paper was about).

Much previous research has broadly agreed that some type of IDENTITY constraint between the elided material (in brackets) and the antecedent material (underlined) licenses the felicitous use of sluicing. However, there is less agreement on what the precise nature of this constraint is, with some researchers defining it over syntactic representations (Sag 1976; Chung et al. 1995; Rudin 2019), some in purely semantic terms (Dalrymple et al. 1991; Merchant 2001; Potsdam 2007), and yet others with reference to both syntactic and semantic levels of representation (Chung 2006; Tanaka 2011a; b; Merchant 2013; Chung 2013; AnderBois 2014).

Historically, data involving syntactic mismatch between the antecedent and ellipsis clauses have been central to such debates, since they provide a diagnostic for the level of representation at which identity constraints hold: Mismatches should render sluicing unacceptable only if it is (at least partially) sensitive to the syntactic representation of the antecedent (Lipták 2015a: inter alia). As we discuss in more detail below, sluicing is unlike certain other cases of ellipsis (most notably, VP-ellipsis) in being widely regarded to disallow syntactic mismatches at the full constructional level, a finding that has strongly shaped the character of contemporary theorizing (Merchant 2013).

The goal of this paper is to expand the empirical base for theorizing about sluicing by examining its felicity under two types of constructional mismatches experimentally: Tough mismatches (Experiment 1) and voice mismatches (Experiment 2). Running counter to the predictions of various contemporary theories, the results reveal a set of highly felicitous sluices with both types of constructional mismatch, as well as a considerable degree of variability across individual items. We explain their ramifications for existing theories of sluicing and ellipsis more generally, arguing that no current movement-based theory captures the entire range of observations. We therefore offer our experimental findings as adequacy criteria to guide future theorizing in the area.

2 Background

The mismatch data scrutinized in the sluicing literature have primarily come in two types. The first type involves mismatches at the lexical level, exemplified in (2).

(2) 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). (modality mismatch; Rudin (2019: ex. 23a))
d. Either the Board grants the license by December 15 or it explains why (the Board didn't grant the license by December 15). (polarity mismatch; Kroll (2019: ex. 30))

Examples (2a)–(2d) all involve mismatches at the level of the verbal complex — finiteness, tense, modality, and polarity respectively — and are nonetheless widely agreed to be felicitous despite the lack of full syntactic identity between the antecedent and elided material (Rudin 2019).

The second type of mismatch is constructional in nature, as shown in (3).

(3) 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)

Examples (3a)–(3d) are mismatches involving the spray/load alternation, the inchoative/transitive alternation, the possessive/accusative gerund alternation, and the active/passive voice alternation respectively. The importance of constructional mismatches has long been recognized (Levin 1982; Merchant 2008; Tanaka 2011b; Merchant 2013; Lipták 2015a; Rudin 2019). Unlike cases of lexical mismatch, however, examples involving constructional mismatch are typically, and perhaps even uncontroversially, judged to be highly unacceptable, as the judgments for (3a)–(3d) would suggest.

Notably, the situation is very different in the VP-ellipsis literature, where constructional mismatches have been central to the debate around identity constraints, and for which the data are notoriously mixed. For instance, whereas the voice mismatch in (4a) is infelicitous, (4b) is widely considered to be at least relatively acceptable.

(4) 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). (Kehler 1993: ex. 3)

(5) a. #It's easy to identify venomous snakes, and poisonous plants are (easy to identify) as well.
b. Venomous snakes are easy to identify, and most experienced hikers can (identify them). (Kertz 2013: 407)

The same goes for the tough-alternation cases (5a)–(5b). 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 (Kertz 2010; 2013; Arregui et al. 2006; Kim et al. 2011; Poppels & Kehler 2019; Kim & Runner 2018). The present paper extends this line of research by bringing experimental work to bear on the mismatch facts around sluicing, focusing on these two constructions as well.

The ensemble of sluicing mismatch facts has received considerable engagement in the literature, forcing an evolution in the thinking of prominent researchers. For instance, whereas Chung et al. (1995) had proposed that sluicing is enabled by a process that copies the syntactic (LF) representation of the antecedent clause into the ellipsis site — thereby giving rise to a syntactic identity condition — Merchant (2001) argued against their proposal due in part to lexical mismatches of the sort shown in (2). He instead proposed a purely semantic identity condition, i.e., e-GIVENness, which merely requires that the existential closures of the denotations of the antecedent and ellipsis clauses entail one another. However, it was soon recognized (Chung 2006) that his approach wrongly allows constructional mismatches such as (3), since e-GIVENness is insensitive to the manner in which propositions are expressed syntactically.

The undergeneration of strictly syntactic approaches and overgeneration of strictly semantic ones led to a series of analyses that propose hybrid identity conditions that reference both semantic and syntactic representations (Chung 2006; 2013; Merchant 2013). For example, Chung (2006) combined e-GIVENness with a lexico-syntactic condition that has come to be known as the NO NEW WORDS constraint (alternatively, "Chung's generalization"), which prevents the ellipsis site from containing any lexical material not provided by the antecedent clause. (Chung (2013) subsequently revised this condition to ban new lexical elements only if they either assign case to the remnant wh-phrase or else determine the argument structure that the remnant participates in.) This condition 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, (5b) can be ruled out under the assumption that passive and active variants of verbs are distinct lexical items (a common assumption since at least Hale & Keyser (1993)), as illustrated in (6) (underlining indicates violations of the No New Words constraint):

(6) Protesters were tear-gassed$_{\textit{passive}}$ but they don't know who #(tear-gassed$_{\textit{active}}$ them).

Rudin (2019) follows a similar logic in ruling out argument-structure mismatches like (3) and (6). Observing that the acceptable mismatches we have seen (2) are all located above the highest elided vP — what he calls the EVENTIVE CORE — he concludes that identity must be restricted to only the material below this node. Hence, whereas the eventive core is required to be syntactically identical to the antecedent, 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 almost all preceding 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. Notably, his identity condition prevents mismatches between lexical items that project different argument structures (small *v* in his case), and further imposes a structure-matching constraint that 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.[1]

It is worth noting that both Chung's (2006) and Rudin's (2019) accounts are explicitly limited to sluicing. And in fact, an attempt to apply them directly 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. Recall that some cases of voice mismatch with VP-ellipsis — e.g. (4b), repeated below as (7) — are relatively acceptable:

(7) This problem should have been solved$_{\textit{passive}}$, but obviously nobody did (solve$_{\textit{active}}$ the problem).

On a unified theory of ellipsis, example (7) should be ruled out on the same grounds as examples (3a)–(3d).[2]

[1] Anand et al. (2023) provide further arguments for Rudin's idea that the identity requirement should be restricted to a subdomain of the elided structure, drawing on evidence from small clause and copular structures. They also propose a minor extension of Rudin's definition of isomorphism to account for pseudo-sluices, but this extension is immaterial with respect to the examples we study here. Since Anand et al. (2023) make the same predictions about those sentences as Rudin's original proposal (Rudin 2019), we will continue to refer to the latter throughout this paper.

[2] Rudin (2019) briefly comments on this issue and concedes that extending his analysis to VP-ellipsis would require walking back his "eventive core" generalization, since allowing voice-mismatched VP-ellipsis would require re-defining the domain of identity as strictly smaller than *vP*.

The challenge of providing a unified explanation for the variable effect of voice mismatches on VP-ellipsis and sluicing prompted Merchant (2008; 2013) to pursue a different approach (see also Tanaka (2011b)). Merchant's (2008) analysis adopts Chung's (2006) hybrid identity account that combines e-GIVENness with the No New Words constraint and, like Chung's, voice mismatches are attributed to a mismatch between active and passive Voice heads.[3] However, 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. Hence, the difference between VP-ellipsis and sluicing derives from differences in the size of the elided constituent and, correspondingly, the domain of identity, as shown in (8) (~~strike-out~~ font is used to indicate the domain of deletion under identity according to Merchant (2013)).

(8) This problem should have been solved, but I don't know…
a. *…who ~~$[_{\text{TP}}$ $[_{\text{VoiceP}}$ solved the problem]]~~.
b. …if anyone ever will $[_{\text{VoiceP}}$ ~~$[_{\text{VP}}$ 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, it has to attribute the fact that some cases, such as (4a), are unacceptable due to independent factors external to the theory of ellipsis.

Thoms (2015) takes a different tack. Arguing against hybrid analyses like Merchant (2013) and Chung (2013), he rejects the notion that syntactic identity is restricted to a subset of "special heads" in the ellipsis clause. He instead advocates for an analysis that includes a Scope Parallelism constraint that requires parallel scope relationships between the antecedent and ellipsis clauses, combined with a mechanism for accommodating syntactic structures to serve as antecedents when the structure that the constraint requires differs from the one present in the antecedent clause. Crucially for our purposes, this antecedent accommodation procedure does not allow the accommodated antecedent to be more syntactically complex than the antecedent clause. Whereas Thoms' account, like Chung's (2013), permits a class of voice mismatches (see Experiment 2), it rules out mismatches due to tough movement, as we explain in greater detail below.

In addition to the movement-based analyses just surveyed, a variety of nontransformational accounts have also been offered (Levin 1982; Ginzburg 1992; Ginzburg & Sag 2000; Jäger 2001; Culicover & Jackendoff 2005; Sag & Nykiel 2011; Barker 2013; Nykiel & Kim 2022: inter alia). Speaking broadly, the constraints on sluicing that such analyses posit are governed by semantic

[3] Murphy (2020) argues that active-voice impersonal constructions in Polish, Irish, and Estonian provide a challenge to Merchant's analysis, on the grounds that it allows for unacceptable mismatches with active-voice sluices. Murphy ultimately proposes, however, that impersonals contain a special type of *v* head, leading to a mismatch that is permitted under Merchant's analysis.

conditions rather than syntactic ones. For this reason, the results of the experiments presented here are largely compatible with the predictions of these theories since the syntactic form of the antecedent is not at issue. As with hybrid approaches, such analyses need to account for the case connectivity effects that appear to render examples such as (3a)–(3d) ungrammatical. We discuss this point in greater detail in §6.

Finally, we note that the accounts surveyed above predict a categorical distribution for sluicing. In light of intuitions regarding examples such as (2) and (3), it is widely assumed in the literature that this is a correct prediction. A goal of this paper is to test this assumption experimentally, employing a broader set of examples than previously considered. Our approach is inspired by the analogous literature on VP-ellipsis, which has found variable mismatch effects for both voice mismatches and mismatches under tough movement (e.g., Kertz (2013)). Experiment 1 considers sluicing under mismatches due to tough movement, and Experiment 2 examines voice mismatches. 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.

3 Experiment 1: Tough mismatches

According to most contemporary syntactic theories (e.g., Messick (2012)), tough movement results in a syntactic trace being left behind after fronting the object, as shown in the first clause of (9).

(9) Banks$_i$ are virtually impossible to rob t_i unless you know when (to rob $\underline{\text{banks}_i}$).

Pairing this structure with an ellipsis clause in which such movement hasn't taken place results in a mismatch between the elided object NP — in (10), *banks* — and the corresponding trace in the antecedent. In contrast to the VP-ellipsis literature, no work of which we are aware has cited felicitous examples involving tough mismatch under sluicing, let alone examined the question experimentally. The goal of Experiment 1 is to do just that.

We begin by surveying the predictions of the accounts introduced in the Background section. Analyses that require full syntactic matching, as represented by the early analysis of Chung et al. (1995), predict tough mismatches to be unacceptable. Specifically, the reconstructed syntactic material will contain a copy of the trace that occurs in the antecedent clause, which will fail to have a licit binder in the ellipsis clause (**unless you know when (to rob* $\underline{t_i}$)).

On the other hand, analyses that rely on a semantic identity condition, most prominently Merchant (2001), straightforwardly predict that such cases will be felicitous. 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, per Merchant's e-GIVENness constraint.

The majority of the hybrid identity theories surveyed in the previous section also predict acceptability, due to various types of flexibility built in to circumvent the undergeneration problems with purely syntactic approaches as discussed in the Background section. For instance, Chung's (2006) lexico-syntactic "No New Words" constraint is unaffected by tough movement since it is by definition insensitive to word order: Whereas 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; 2014; Merchant 2013) – predict that tough movement should not affect the acceptability of sluicing.

Similarly, 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. As detailed below, our experimental materials exclusively feature *how*, *when*, and *where* remnants that render them exempt from this limited syntactic identity condition. Sluicing is therefore predicted to be possible.

Rudin's (2019) syntactic identity condition is more restrictive than the other hybrid theories mentioned so far, but still permits sluicing under tough movement. Whereas his account requires structure matching in addition to lexical identity, which in turn requires the elided object *banks* to be compared to the trace in the antecedent clause instead of the fronted NP, his 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:

(10) 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, Rudin (2019) also derives the acceptability of sluicing under tough movement. In fact, accounts that adopt the vehicle change provision need not consider tough mismatches to be a true case of constructional mismatch, since the structure required at the ellipsis site only differs from the antecedent with respect to the trace/binder alternation.[4]

[4] Another type of mismatch that falls in this category involves alternation between topicalized and non-topicalized sentences, as in (i).

Finally, the analysis of Thoms (2015) predicts tough mismatches to be unacceptable. Whereas his Scope Parallelism constraint has similar consequences as Rudin's (2019) structure-matching condition with respect to word order, unlike Rudin, Thoms explicitly prohibits lexically distinct items to count as identical if they are syntactically co-indexed. 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.[5] As such, Thoms' analysis rules out mismatches due to tough movement.

Against this theoretical landscape, our aims in conducting Experiment 1 are threefold. First, in bringing tough-mismatch examples to the fore, we provide the first experimental evaluation of their acceptability status. Second, we use the results to assess the foregoing analyses and their varying predictions, as well as inform the development of new theories. Finally, we ask whether the results point to a categorical acceptability status, or whether they give rise to gradient effects that would provide a new type of adequacy criterion for current and future accounts.

3.1 Methods

3.1.1 Stimuli

24 items were constructed, each with 12 variants according to a 2 × 2 × 3 within-item and within-participant design, as shown in (11).

(11) 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]

(i) Banks$_i$, you shouldn't rob t_i unless you know when (to rob banks$_i$).

It is important to stress, however, that appealing to vehicle change to account for such cases comes at a theoretical cost, as it is not an explanatory account of mismatches in referential form, but instead a descriptive constraint stipulated in lieu of an explanatory account. We therefore agree with Merchant (2001: 25) when he says:

> To pursue a theory of ellipsis based on structural isomorphism 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.

Note that syntactic analyses would have had no problem accounting for cases like (9), (10), or (i) if they had turned out to be unacceptable, by simply declaring that traces do not participate in vehicle change.

[5] 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.

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]

Independent manipulations included the presence or absence of ELLIPSIS, whether or not there was a MISMATCH, and which of three WH-WORDS were employed, none of which are associated with argument positions (*how*, *when*, *where*). Unelided variants were included to ensure that any penalties witnessed in the ellipsis conditions were in fact due to ellipsis and not independent factors. Further, 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 (Kertz 2013; Gordon et al. 1993). To avoid 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).

While the within-item manipulation of WH-WORD has the benefit of increasing the lexical variety of the experimental materials, it does creates a potential issue with respect to plausibility, since all three questions may not be equally plausible in the same context. For example, whereas a *how* question is clearly a plausible follow-on to each of the first clauses in (11), 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 variants and thus will not interfere with the interpretation of our results against our central questions regarding 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 2 × 2 × 3 within-item manipulation, the matrix clause of the embedded question was varied across experimental items in the manner shown in **Table 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 filler items were non-elliptical, and both elliptical and non-elliptical fillers included acceptable and unacceptable sentences, as exemplified in (12).

(12) 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]

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

Table 1: Range of sluice-embedding clauses across experimental items.

3.1.2 Participants and procedure

43 participants were recruited via Amazon.com's crowd-souring platform Mechanical Turk.[6] Using the Ibex platform for web-based experiments (Drummond 2017), 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 558 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, of which 816 corresponded to experimental items and were analyzed as follows.[7]

[6] We aimed for a minimum of 40 participants for each experiment, but due to the nature of the online recruitment process we ended up with slightly more participants even after applying exclusion criteria (e.g., removing participants who identified as non-native speakers of English).

[7] The data and code needed to reproduce all statistical analyses and visualizations reported in this paper are available at https://github.com/tpoppels/poppels-kehler-sluicing-mismatch-paper.

3.2 Results

Population-level averages are shown in **Figure 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. Second, 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.

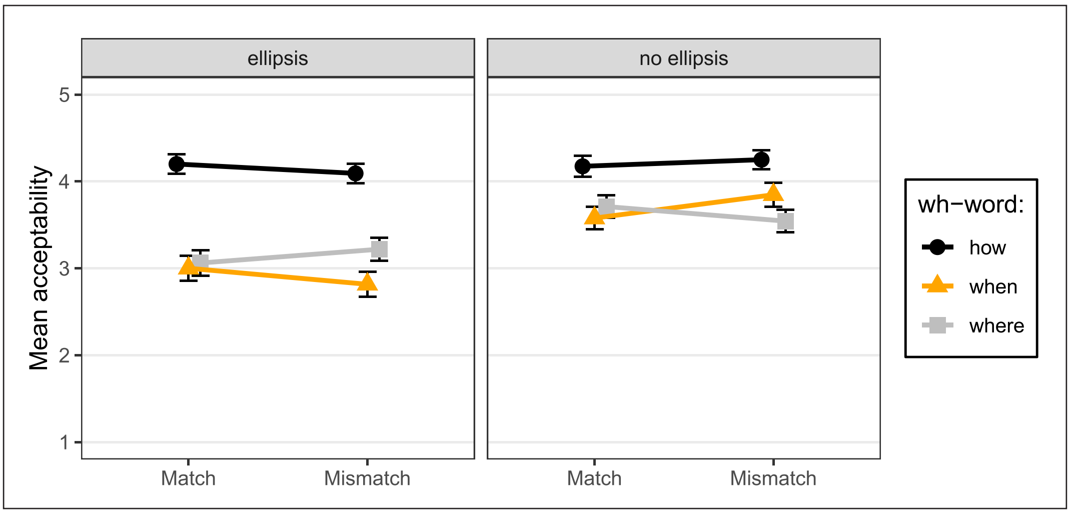


Figure 1: Condition averages from Experiment 1. The average acceptability of filler items ranged from 1.7 to 4.8.

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 2 × 2 × 3 design (including all interactions) and crossed group-level effects for items and participants including individual intercepts and slopes for all population-level effects. Both analyses were conducted with brms, an R package for Bayesian multilevel models (R Core Team 2021; Bürkner 2017; 2018).[8] The first model was designed to test our primary hypothesis, i.e.

[8] The formula used for both models was response ∼ ellipsis*mismatch*wh.word + (1 + ellipsis*mismatch*wh.word | subject) + (1 + ellipsis*mismatch*wh.word | item). The models were fit with weakly informative priors on all parameters according to a normal distribution with a standard deviation of 4. We sampled from 4 chains and 4000 iterations in total, 1000 of which were warm-up samples to prevent any effect of initialization (as is typical for probit models, the parameters were initialized at 0). In response to a question from an anonymous reviewer, we emphasize that not all reported effects directly correspond to a model parameter, which is why we report them using Δ rather than β, along with the Credible Interval $CI(\Delta)$ around Δ and the posterior model probability that Δ is above or below 0 (depending on the question at hand).

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.[9] This analysis revealed 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 item-specific effects were modeled with “shrinkage” towards the population-level effects,[10] we were able to explore whether the *when/where* penalties differed across items without manually adjusting for multiple comparisons (Gelman et al. 2012). **Figure 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.

[9] Note that our primary hypothesis test boils down to a 2 × 2 comparison across MISMATCH and ELLIPSIS. As the remaining 3-way factor WH-WORD is irrelevant to the question it addresses, it was sum-coded, which lends additional statistical power to the hypothesis test.

[10] 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 are more conservative compared to non-hierarchical models, which permits multiple hypothesis tests without inflated significance thresholds.

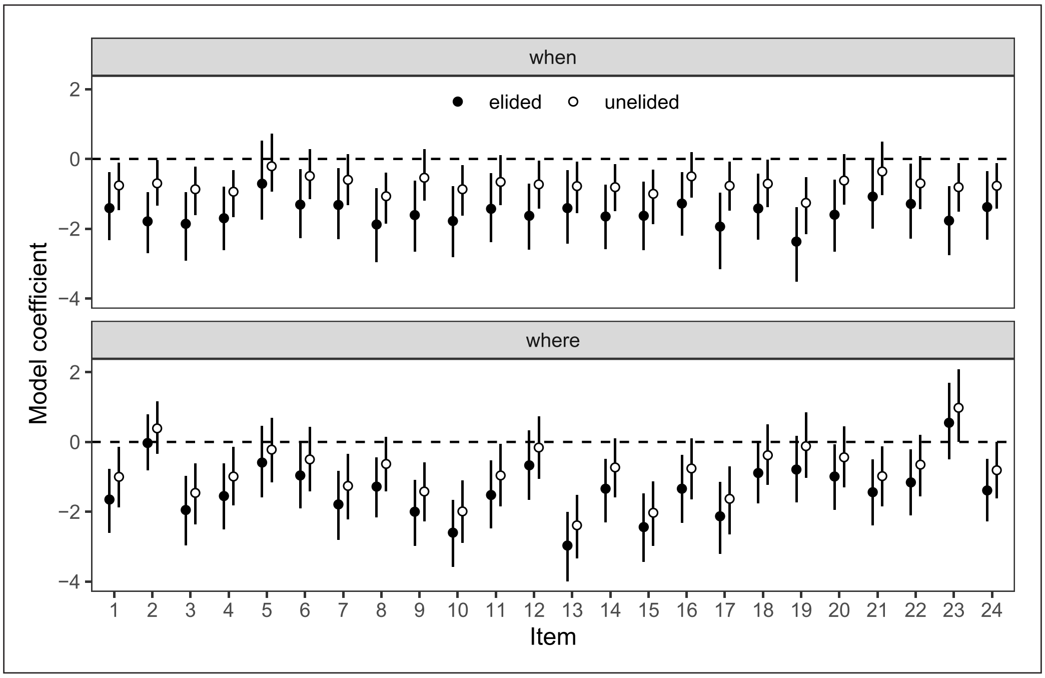


Figure 2: Item-specific model coefficients corresponding to the penalties associated with *when* (top) and *where* (bottom) questions relative to the *how* variant of the same item. Black dots indicate elided variants; white dots represent unelided variants. Errorbars show 95% Credible Intervals (i.e., 95% of the posterior samples fall within that interval).

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 smallest *when/where* degradation and the three that exhibited the largest. Consider first the *when* questions in (13), which showed the least evidence for a penalty relative to their respective *how* counterparts.[11]

(13) 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. These judgments contrast strongly with those for the three *when* items that exhibited the greatest penalty relative to their *how* counterparts:

[11] While we list mismatch variants here, note that the items were selected based on the model coefficients shown in **Figure 2**, which represent the across-the-board difference between *when/where* variants and their *how* counterparts, i.e. averaging over matched and mismatched variants.

(14) 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)
c. Pants that fit perfectly can be impossible to find unless you know when. (Item 19)

Intuitively, the embedded questions are less plausible continuations in (14) than in (13), suggesting that the effect in question may be driven by question plausibility.

The *where* questions that revealed the least evidence for a penalty relative to their respective *how* counterparts are shown in (15). Item 19 is particularly informative because its *when* variant — shown in (14c) — was significantly degraded whereas its *where* variant — shown in (15a) — was not.

(15) 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)

Indeed, none of these three items were degraded, which is again consistent with the analysis that the *when/where* degradations are driven by question plausibility. In fact, Item 23, shown in (15c), exhibits the opposite effect, such that the *where* variant was more acceptable than the *how* variant. This is consistent with the fact that knowing *where* to park large cars is presumably a more plausible bottleneck to parking them downtown than knowing *how* to do so.

Finally, the items exhibiting the largest *where* penalty are shown in (16).

(16) 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)

Once again there's a sense in which question plausibility is relevant. 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 3**, there was indeed a high correlation between the two, especially for *where* items (*when*: $r = 0.85$; *where*: $r = 0.995$).

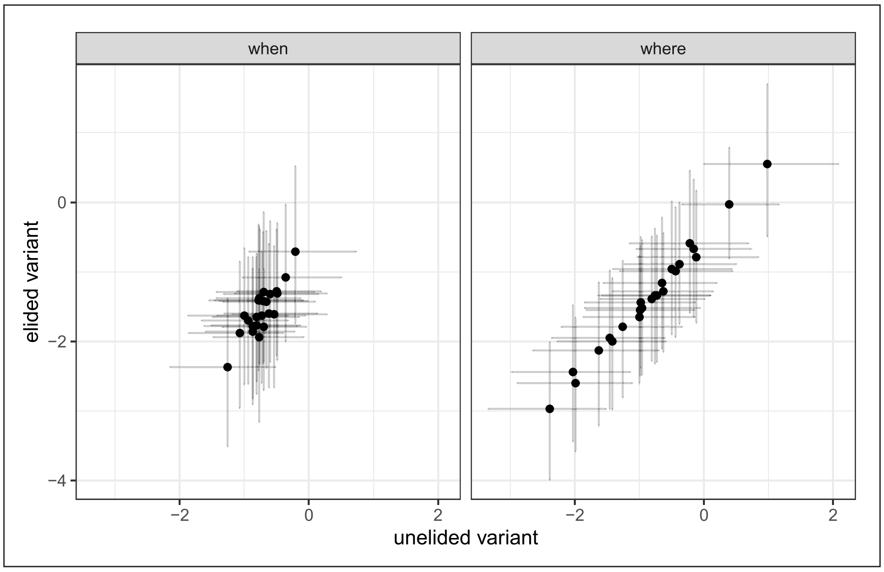


Figure 3: Scatterplots showing correlation between item-specific *when/where* penalties associated with unelided (x) and sluiced (y) variants. Errorbars show Credible Intervals around model coefficients.

3.3 Discussion

The key findings from Experiment 1 are twofold. First, we found no evidence that mismatch due to the *tough*-alternation negatively affects the acceptability of sluicing. Second, *when* and *where* sluices were significantly degraded compared to their *how* counterparts, a finding that may reflect a penalty for implausible questions that is exacerbated 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 movement-based theories of sluicing. As mentioned in this section's introduction, the common wisdom is that tough movement leaves behind a syntactic trace after the object is fronted (e.g., Messick (2012)), resulting in a mismatch since the elided object NP — *banks* in (9), repeated below as (17)— corresponds to a trace in the antecedent clause.

(17) Banks$_i$ are virtually impossible to rob t_i unless you know when (to rob banks$_i$).

For this reason, the early analysis of Chung et al. (1995), fails to predict the acceptability of such cases, since the reconstructed trace in the ellipsis clause will fail to have a licit binder.

The results support the predictions of the majority of other theories surveyed, however. Purely semantic identity accounts that appeal to Merchant's (2001) e-GIVENness condition predict them straightforwardly, since tough movement does not affect the mutual-entailment relation between the antecedent and elided material. Chung's (2006) analysis similarly makes the right predictions, since her No New Words constraint is unaffected by tough movement. Likewise, since neither the case-assignment condition nor the argument-structure condition posited by Chung's (2013) account apply to tough movement, the results are captured by this analysis as well. The results are further consistent with Rudin's (2019) purely syntactic identity condition, in light of its appeal to vehicle change in treating traces as interchangeable with their binders.

Finally, recall that Thom's (2015) analysis is based on a Scope Parallelism requirement that is violated by tough movement, since his account explicitly prohibits lexically distinct items to count as identical if they are syntactically co-indexed. The results of Experiment 1 are therefore problematic for his proposal.

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); Merchant (2008; 2013) – 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 closer look at voice mismatches.

4 Experiment 2: Voice mismatches

As we outlined in the Introduction, mismatches under sluicing have been widely assumed to be ungrammatical, in light of unacceptability of examples with active antecedent clauses and passive ellipsis clauses like (3d), repeated below as (18a), as well as cases with passive antecedent clauses and active ellipsis clauses such as (18b).

(18) a. #Someone abducted the candidate, but we don't know by who (the candidate was abducted). (Chung et al. 2011: ex. 25b)
b. #The candidate was abducted, but we don't know who (abducted the candidate).

Intuitive judgments clearly indicate that (18a)–(18b) are unacceptable, hence why the literature has taken the prediction of such unacceptability to be a determining factor for the adequacy of analyses. To our knowledge, however, previous authors have considered only examples in which the remnant corresponds to an argument position of the verb in the antecedent, as is the case in (18a)–(18b). In contrast, in Experiment 2 we consider voice mismatches in which the remnant corresponds to an adjunct position, as in (19).

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

The analyses of Chung et al. (1995); Chung (2006); Merchant (2008; 2013); Rudin (2019) predict that sluicing is ungrammatical under voice mismatches across-the-board, attributing the cause to whatever lexical items encode Voice. Chung (2013) and Thoms (2015), on the other hand, each allow for (19) while still ruling out (18a)–(18b). Specifically, according to Chung's (2013) "limited syntactic identity" account, only non-identical heads that either take the sluicing remnants as an argument or assign Case to a remnant DP are prohibited. Neither of these is the case for the adverbial remnants in our experimental materials, unlike (18a)–(18b).

Thoms' (2015) Scope Parallelism account also predicts the absence of a mismatch penalty, although for different reasons. Whereas he rejects Chung's restriction of the identity condition to "special heads" on theoretical grounds, his analysis requires that the antecedent and ellipsis clauses exhibit parallel scope relations. As a result, voice-mismatched sluices with correlates in argument position, such as (18), are ruled out on the grounds that the sluicing remnant (in this case, *who*) is extracted from a non-parallel position to its correlate (*someone*). With respect to the *how*, *when*, and *where* items we use Experiment 2, however, the correlate of the wh-phrase is not an argument of the antecedent verb (and is, indeed, implicit). The antecedent and ellipsis clauses therefore exhibit parallel scoping relations, and sluicing is correctly predicted to be acceptable.[12]

Because sluicing remnants are minimal compared to the remnants in VP-ellipsis, one faces an immediate challenge when attempting to construct mismatched cases in which the antecedent is passive and ellipsis clause is active. Specifically, any attempt to evaluate the acceptability of the mismatch in a case like (20a) is confounded by the possibility that an addressee will instead recover a voice-matched interpretation as shown in (20b).[13]

(20) The problem has finally been solved but I don't know…
 a. …how (they solved it). [voice-*mis*matched]
 b. …how (it has been solved). [voice-matched]

The problem isn't insuperable, however. In this experiment, we take two steps to avoid the issue. First, we design the materials in a way that renders the syntactically-matched, passive-voice interpretation of the ellipsis clause implausible. Consider (21).

(21) The problem hasn't been solved because no one knows…
 a. …how #(the problem hasn't been solved).
 b. …how (to solve it).

[12] Thoms concedes that his analysis of acceptable cases of sprouting, i.e. sluicing without overt correlates, requires the assumption that implicit correlates are syntactically represented in the antecedent and can therefore participate in scope relations that are parallel to the ellipsis clause. With this assumption in place, his analysis correctly predicts that sluices involving *how*, *when*, or *where* remnants are unaffected by voice mismatches, while maintaining the prediction that argument-targeting sluices become unacceptable.

[13] See Chung (2013) for a similar point.

Whereas the ellipsis clause can in principle be interpreted in a way that preserves syntactic identity (per (21a)), doing so would lead to an implausible construal 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*, per (21b).

Second, in order to not merely rely on our intuitions about such cases, we conducted a separate norming experiment to confirm which interpretations participants did, in fact, adopt. This experiment is described in the section entitled "Norming experiment" below.

4.1 Methods

4.1.1 Stimuli

We created 12 experimental items that followed a 2 × 2 × 3 design that independently varied the presence/absence of ELLIPSIS and MISMATCH within items, as shown in (22).

(22) 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 ever 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.[14]

In addition to the 12 experimental items, participants were presented with 12 filler items designed for a separate experiment. As in Experiment 1, they included both elliptical and non-elliptical sentences and covered the range of the acceptability scale. Four representative examples are shown in (23).

(23) a. Sarah is jealous but she didn't say what. [+ellipsis, –acceptable]
b. The package was delivered somewhere, but no one seems to know where. [+ellipsis, +acceptable]
c. The customer left the store, but it is unclear who left. [–ellipsis, –acceptable]
d. The concierge was reading the newspaper, but I couldn't see which one he was reading. [–ellipsis, +acceptable]

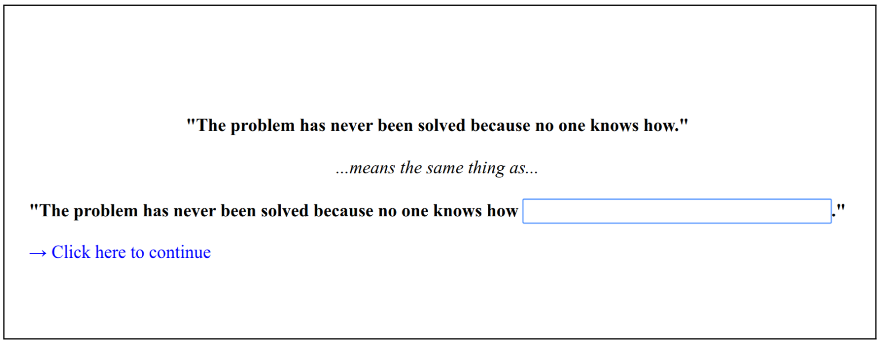
[14] An anonymous reviewer rightly points out that the number of items for experiment 2 was lower than seen in typical acceptability judgment experiments. We intentionally limited the included items to high-quality items to ensure the naturalness and plausibility of the sluiced questions.

4.1.2 Participants and procedure

We recruited a total of 52 participants from Amazon.com's Mechanical Turk platform. The procedure was identical to that of Experiment 1: Participants judged the acceptability of experimental items and fillers on a scale from 1 ("unacceptable") to 5 ("acceptable") via the Ibex platform for web-based experiments (Drummond 2017). 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 1000 ms, leaving us with a total of 930 observations from 47 participants, of which 469 data points corresponded to experimental items and were analyzed as detailed below.

4.1.3 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 in a norming experiment. 13 participants either reported being non-native speakers or submitted clearly bot-like responses[15] and were excluded from the analysis. Also excluded were 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 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.



"The problem has never been solved because no one knows how."

...means the same thing as...

"The problem has never been solved because no one knows how ________**."**

Figure 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.

[15] 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."

This conclusion, of course, assumes that the paraphrase task provides reliable evidence of the participants' interpretation of the ellipsis clause. Here we follow Frazier & Duff (2019) who, in pursing a syntactic reconstruction account, argue that comprehenders' paraphrases of elliptical utterances are likely to re-use the syntactic material they infer when resolving the ellipsis. But it should be kept in mind that 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 (24a), or else have to contend with additional lexical mismatches, as exemplified in (24b).[16]

(24) 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).

Indeed, all eight of the passive-voice paraphrases in the norming experiment introduced a modal mismatch as shown in (24b), and further seven of those also introduced a voice mismatch (i.e., these paraphrases were produced in response to an active-voice antecedent). In other words, only a single response to a MISMATCH item (0.7%) employed a modal mismatch in order to avoid a voice mismatch. While it is in principle possible that the passive-active mismatches in our experiment were felicitous because they were understood as passive-passive sluices with a mismatching modal, one would expect those interpretations to show up more frequently in comprehenders' paraphrases of the ellipsis site. We are thus confident that the MISMATCH items presented to participants in Experiment 2 reflect genuine voice-mismatched sluices.

4.2 Results

Population-level averages are shown in **Figure 5**, in which two patterns emerge visually. First, there does not appear to be a robust mismatch penalty (horizontal lines are not consistently downward-sloping). Second, 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 2 × 2 design of the experiment, with sum-coding for both ELLIPSIS and MISMATCH (so that main effects can be interpreted as "across-the-board" effects) and added WH-PHRASE as a grouping factor alongside items and participants.[17] This

[16] While Rudin (2019) explicitly permits mismatches above the highest elided small *vP* (the "eventive core"), (24b) does violate less forgiving identity conditions, including those adopted by Chung (2006; 2013); Merchant (2013; 2008).

[17] As in Experiment 1, all analyses were conducted with the brms R package for Bayesian multilevel models (R Core Team 2021; Bürkner 2017; 2018), with weakly informative priors on all parameters according to a normal distribution with a standard deviation of 4. As usual, all group-level intercepts and slopes corresponding to the 2 × 2 population-level effect structure were added (Barr et al. 2013). The model formula was: response ~ ellipsis*mismatch + (1 + ellipsis*mismatch | subject) + (1 + ellipsis*mismatch | item) + (1 + ellipsis*mismatch | wh.word). We sampled from 4 chains and 4000 iterations in total, 1000 of which were warm-up samples to prevent any effect of initialization (as is typical for probit models, the parameters were initialized at 0). In response to an anonymous reviewer, we remind the reader here that not all reported effects directly correspond to a model parameter, which

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). Second, 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).

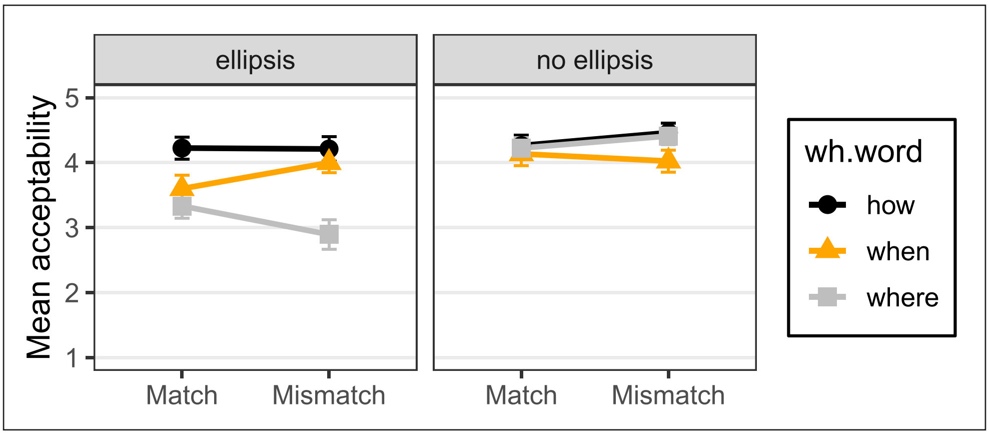


Figure 5: Condition averages from Experiment 2. The average acceptability of filler items ranged from 1.4 to 5.0.

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 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*: $\Delta = -0.19$, $CI(\Delta)$

is why we report them using Δ, along with the Credible Interval $CI(\Delta)$ and the posterior model probability that Δ is above or below 0 (depending on the question at hand).

$= [-0.81, 0.37]$, $P(\Delta < 0) = 0.75$). 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*: $\Delta = -0.68$, $CI(\Delta) = [-1.66, 0.31]$, $P(\Delta < 0) = 0.91$).

4.3 Discussion

The key result from Experiment 2 is that no evidence for a mismatch penalty was found: 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) and Thoms (2015), to which we return momentarily): Chung (2006); Merchant (2008; 2013); 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 (2013) and Rudin's (2019) systems, and that the passivization of the antecedent clause additionally incurs a violation of Rudin's (2019) structure-matching constraint.

As described earlier, Chung's (2013) "limited syntactic identity" condition is consistent with our results, since only non-identical heads that either take the sluicing remnants as an argument or assign Case to a remnant DP are prohibited, neither of which is the case in our experimental materials. Thoms' (2015) Scope Parallelism account likewise predicts the absence of a mismatch penalty, since in our materials the correlate of the wh-phrase is not an argument of the antecedent verb. As a result, the antecedent and ellipsis clauses therefore exhibit parallel scoping relations, and sluicing is correctly predicted to be acceptable.

5 A Note on Tanaka (2011a)

Tanaka (2011a) claims that examples of the sort we have utilized in our two experiments are not actually instances of sluicing at all. Instead, he argues that they represent a combination of two other types of ellipsis: VP-ellipsis followed by an independent rule that elides the infinitival *to*. So whereas the surface form of example (25a) has all of the morphosyntactic hallmarks of sluicing, Tanaka suggests that the correct analysis is actually as shown in (25b), in which two distinct ellipsis processes have applied.

(25) a. I'll fix the car if you tell me how (~~to fix the car~~).
b. I'll fix the car if you tell me how (~~to~~) (~~fix the car~~).

Tanaka's concern is not the existence of structural mismatch, but instead the fact that the infinitival *to* in (25a) can be deleted at the ellipsis site even though it doesn't occur in the antecedent — a violation of syntactic isomorphism and the No New Words constraint. Rather than taking such examples to present problems for these constraints (as Merchant (2001) does with respect to syntactic isomorphism), Tanaka attempts to rescue the constraints by arguing that these examples are not actually instances of sluicing. If Tanaka were shown to be right, then the experiments we have presented would not concern sluicing at all.

We do not find Tanaka's proposal to be convincing, however. Indeed, we see four problems with it: (i) it is stipulative, and hence unmotivated and unexplanatory, (ii) it is not consonant with our experimental results nor can it be made to be, (iii) it makes incorrect predictions regarding other cases, and (iv) certain arguments he presents don't fully go through, or otherwise have alternative explanations. We expand on these in turn.

First, Tanaka's posited rule receives no independent motivation. In considering its plausibility, the first thing one would want to see is evidence of the existence of a rule for infinitival deletion elsewhere in the grammar, but none is presented. Without such evidence, it is hard to see the constraint as more than a stipulation designed to dispense with a troublesome fact when trying to salvage the adequacy of a proposed set of constraints, constraints for which there a variety of other problematic cases in the literature. The stipulatory nature of the rule in fact goes further, in that it is only posited to apply after *how*, and not other types of remnants: "ellipsis of 'to' is an idiosyncratic property of *how*, which should not be generalized beyond this particular lexical item" (Tanaka 2011a: 90). Again, no independent evidence is provided for why such a strangely idiosyncratic rule should exist.

Second, Tanaka's proposal is directly countered by our experimental results, and hence is empirically inadequate: we have presented a variety of cases in which sluices with mismatched antecedents involving *when* and *where* remnants are judged to be highly felicitous. The following sentence, for example, received an average rating of 4.0 out of 5 and is statistically indistinguishable from the acceptable fillers in the experiment:

(26) I can't believe Joe still hasn't been fired just because his boss can't decide on when (to fire him).

In order to account for this fact under Tanaka's analysis, his *to* deletion rule must be extended to other adverbial remnants beyond *how*. However, as Tanaka himself notes, any attempt at generalizing the rule in such a way is doomed to fail. This is particularly apparent in cases involving *whether* remnants, which are famously disallowed in English sluicing, as shown in (27a):

(27) a. John is allowed to go to the meeting, but he doesn't know whether #(to go to the meeting).

b. John is allowed to go to the meeting, but he doesn't know whether to (go to the meeting).

Note that whereas the sluicing in (27a) is strongly unacceptable, the VP-ellipsis in (27b) is perfectly felicitous. But if (27b) were a possible source for (27a), to which an infinitival deletion operation could then apply, one would expect (27a) to be as acceptable as (27b). That is, even if a constraint against *whether* remnants in sluicing exists, an infinitival deletion rule should provide an alternate path of acceptability for (27a) that doesn't involve sluicing. Again, a mere statement that his rule doesn't apply after *whether* has no explanatory force, as it merely describes the data. Either way one slices it, Tanaka's rule is not only stipulative, but either undergenerates (if idiosyncractically constrained to *how*) or overgenerates (if applied to all other adverbial remnants, as one would expect such a rule to do).

Third, even if one were to stipulate that Tanaka's *to* deletion rule applies to *how*, *when*, and *where* remnants, but not to *whether*, it is nonetheless vulnerable to counterexamples. Consider the following examples of antecedent-contained deletion, a phenomenon that is generally disallowed in English sluicing unless the ellipsis site is embedded inside an adjunct PP (Lipták 2015b), which is not the case here:

(28) a. John only managed to fix the problems that Bill told him how to (fix).
b. John only managed to fix the problems that Bill told him how #(to) #(fix).

Unsurprisingly, (28a) is perfectly felicitous: it involves VP-ellipsis, which is known to allow antecedent-contained deletion. If Tanaka's proposal is correct, however, (28b) should be equally felicitous since it can be derived from (28a) by simply applying his *to* deletion rule. However, it is infelicitous, which suggests that it does involve sluicing after all. In other words, Tanaka's rule cannot explain the unacceptability of (28b) in light of the fact that its putative source, (28a), is perfectly felicitous.

Finally, the primary evidence Tanaka offers for his analysis involves cases in which the sluiced material does not contain a negation, despite the occurrence of one in the putative antecedent. Consider (29), from Merchant (2001: 22).

(29) I can't play quarterback: I don't even know how (to play quarterback).

The most salient reading for (29) is not that the speaker "doesn't know how they can't play quarterback" — as one would expect if the meaning of the sluiced material were constrained to be that of the entire first clause — but instead that the speaker "doesn't know how to play quarterback": the meaning of the ellipsis clause does not involve negation even though its antecedent does.

After analyzing (29a) as VP-ellipsis plus *to* deletion, rather than sluicing, the polarity mismatch now resides outside the ellipsis site, thus no longer running counter to the structural isomorphism requirement:

(30) I can't play quarterback: I don't even know how (to) (play quarterback).

However, the problem of polarity reversals has since been shown to be more far-reaching, including examples like (31) from Kroll (2019: 27; exs. 29–30), which cannot be explained away in the same fashion:

(31) a. Either the Board grants the license by December 15 or it explains why (the Board did not grant the license by December 15). = (2d)
b. Either John didn't do an extra credit problem, or he didn't mark which one (he did do).

Both of these examples contain a polarity mismatch between the elided material and its antecedent, but neither of them can be rescued through Tanaka's analysis since they cannot be construed as VP-ellipsis plus *to* deletion. Lest one worry that Kroll's polarity reversal examples are idiosyncratically linked to *why* and *which one* remnants, in which case Tanaka's analysis may still be relevant to the *how*, *when*, and *where* adverbials at stake in this paper, we note that it is straightforward to construct analogous examples for those remnants:

(32) a. Either John really didn't find out, or he is embarrassed to admit how (he found out).
b. Either Susan didn't do it, or she lied about when (she did it).
c. Either it's true that you never buried the treasure, or you simply don't want to tell us where (you buried it).

Since the attested interpretations do not involve infinitival *to*, the polarity reversal in these examples cannot be explained away as VP-ellipsis and *to* deletion. As such, they undermine what Tanaka himself considers the strongest evidence for his analysis: its ability to rescue structural isomorphism accounts of sluicing from the challenge posed by polarity mismatches.[18]

Tanaka's primary motivation for hypothesizing his rule is to maintain the syntactic isomorphism requirement that his theory posits. In this regard, it is worth noting that the isomorphism assumption faces a variety of other challenges beyond polarity mismatches that likewise cannot be explained through his analysis. Experimental studies reported on by Poppels & Kehler (2023),

[18] Yoshida (2010) also addresses cases in which the meaning of negation (and modals) is not inherited from the antecedent, such as (i).

(i) John isn't inviting anyone without saying who (he <u>is</u> inviting).

Based on this behavior and others, Yoshida argues that the preceding VP is the antecedent and not the entire clause. Yoshida still considers such cases to be sluicing, however, and indeed Tanaka's strategy for viewing apparent cases of sluicing as VP-ellipsis is inapplicable here.

for instance, have identified a wide variety of cases that demonstrate a dissociation between a sluice and its antecedent not only in terms of syntactic isomorphism, but also their meanings. Their stimili included examples like (33)-(36).

(33) a. A: Can I get a few autographs?
b. B: Sure, how many (do you want)?

(34) a. A: Can I borrow your textbook?
b. B: Which textbook (do you want/need to borrow)?
c. B: Why (do you want to borrow it)?

(35) a. A: Did you not tell your friends about the game today?
b. B: I did, but I forgot to tell them where (it would take place).

(36) Regarding Trump's impeachment, the only question is when (he will be impeached).

In (33), B isn't asking *How many autographs can you get (a few of)?*, but a question akin to *How many do you want?*. In B's responses to A in (34b)-(34c), B asks questions along the lines of *Which textbook do you want to borrow?* and *Why do you want to borrow it?*, rather than ones with meanings that match A's question (*Which textbook can you borrow?* and *Why can you borrow my textbook?* respectively). Similarly in (35), the interpretation of B's response is roughly *where it would take place*, and not *where I did tell my friends about the game today*. And finally, the meaning of the ellipsis clause in (36) can be paraphrased as *when he will be impeached*; in this case there is no clause to serve as an antecedent. In all of these cases, experimental participants were happy to interpret those sluices inferentially in a way that goes beyond the antecedent-provided meaning – with interpretations that violate syntactic isomorphism, e-GIVENness, and the No New Words constraint – and yet they nonetheless rate them as highly acceptable.

To conclude, it has become an unfortunate trend in theorizing about ellipsis, in the face of counterexamples to deeply-held morphosyntactic constraints (syntactic isomorphism, No New Words), to salvage those constraints by 'fine-tuning' one's analysis with idiosyncratic rules that lack independent justification. Whereas Tanaka provides certain arguments for his claim that the cases under scrutiny, despite appearances, involve VP-ellipsis, their convincingness falls well short of the level necessary to compel the adoption of the surprising and otherwise unmotivated rule that he proposes. Further, a previously unnoticed problem for the account (antecedent-contained deletion) remains. As we have seen, the examples examined in our experiments are just of one type of case that casts doubt on the existence of a syntactic isomorphism constraint; further, the gradience we have identified in the data is at odds with any approach capable of making only categorical predictions. In sum, the foregoing arguments cast enough doubt on Tanaka's proposal to render it far too premature to assume that the cases under scrutiny — which have all of the hallmarks of sluicing — are actually derived by an alternative process and hence irrelevant to the theory of sluicing.

6 General Discussion

The central 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. For mismatches under tough movement, the results run counter to the predictions of only the early analysis of Chung et al. (1995) and more recent account of Thoms (2015); the other analyses we have surveyed correctly predict felicity. The relative acceptability of the voice mismatches examined in Experiment 2, however, is consistent only with Chung (2013), Thoms (2015), and the nontransformational analyses that we cited in §2 and discuss in greater detail below. 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; 2013; Tanaka 2011b): Despite the fact that VoiceP is elided in our experimental materials, it can nonetheless deviate from its correlate in the antecedent without incurring an acceptability penalty. **Table 2** provides a scorecard for syntactic approaches.

Finding	**Chung et al. (1995)**	**Chung (2006)**	**Merchant (2013)**	**Chung (2013)**	**Thoms (2015)**	**Rudin (2019)**
Acceptable tough mismatches (Expt. 1)	✗	✓	✓	✓	✗	✓
Acceptable voice mismatches (Expt. 2)	✗	✗	✗	✓	✓	✗
when/where penalty (Expts. 1 & 2)	✗	✗	✗	✗	✗	✗

Table 2: Cross-tabulation of empirical findings and theories of sluicing that require syntactic identity.

The only movement-based account that is consistent with the absence of a mismatch penalty revealed by both of our studies is Chung (2013). However, despite being much less restrictive than the other syntactic identity accounts we considered, Thoms (2015) points out that Chung's (2013) account is nonetheless too restrictive with respect to examples like (37).

(37) I remember someone complaining, but I can't remember who (complained).

Thoms (2015) notes that this example appears to be fully acceptable despite the fact that it fails Chung's (2013) 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 shortcoming (along with considerations of theoretical parsimony), Thoms (2015) rejects the notion that syntactic identity is restricted to a subset of "special heads" in the ellipsis clause, and instead advocates for his Scope Parallelism requirement. As we have noted, however, while his account, like Chung's (2013), 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 (2019) allows lexically distinct items to count as "identical" if they are syntactically co-indexed, Thoms (2015) explicitly prohibits such equivalencies. We are thus left with a situation where no movement-based account captures the full range of mismatch patterns under consideration: Whereas Thoms (2015) improves on Chung (2013) with respect to cases like (37), he incorrectly rules out the tough movement cases from Experiment 1.[19]

As mentioned in §2, the results of our experiments are largely consistent with a variety of nontransformational analyses that have been proposed (Levin 1982; Ginzburg 1992; Ginzburg & Sag 2000; Jäger 2001; Culicover & Jackendoff 2005; Sag & Nykiel 2011; Nykiel & Kim 2022: inter alia). Although they vary in their details, all posit that no unpronounced syntactic material exists at sluicing sites, and hence there are no constraints that apply to such structure. Instead, sluicing interpretation is governed by a referential process that identifies an antecedent that serves up a suitable proposition as the target for interpretation.[20] As such, neither the tough

[19] We note that several recent works have posited the potential acceptability of certain types of mismatches as a way to deal with the apparent insensitivity of sluicing to island violations (see Abels (2018) for a review). Consider (i) (Merchant 2001; Barros et al. 2014: inter alia):

(i) a. They hired someone who speaks a Balkan language – guess which!
b. *They hired someone who speaks a Balkan language – guess which they hired someone who speaks!
c. They hired someone who speaks a Balkan language – guess which it is *t*?

As Ross (1969) famously noted, example (ia) is perfectly grammatical, despite the fact that the corresponding unelided version in (ib) is unacceptable due to an island violation. This is a problem for theories that posit the existence of syntax at the ellipsis site, since (ia) is predicted to have the same structure as (ib), and hence should share the same negative grammaticality status. One proposed remedy, assuming a semantic identity condition such as Merchant's, is that the insensitivity to islands might be due to the possibility of alternate possible structures existing at the ellipsis site, such as the cleft shown in (ic). Important details of this proposal have yet to be worked out, such as what the limits are on possible material at the ellipsis site (the "too many paraphrases" problem), how unacceptable cases of mismatch can still be ruled out, and how the hearer can come to identify the missing syntactic material (Abels 2018).

[20] Strictly speaking, Levin's 1982 LFG analysis involves reconstruction at the level of F-structure, but shares with other approaches the lack of a role for surface (C-)structure.

mismatches explored in Experiment 1 nor the voice mismatches studied in Experiment 2 are problematic for these analyses.

Since no silent structure exists at the ellipsis site for constraints to apply to, nontransformational approaches necessarily appeal to other means to account for the case connectivity effects witnessed in (3a)–(3d). The analyses of Jäger (2001) and Barker (2013), for instance, posit that the lexicon is equipped with silent proforms for each possible case; case connectivity results from the appropriate form having to agree with the case of both the antecedent and the remnant. Whereas this move may initially strike some as being stipulative, proponents of nontransformational approaches have in fact provided independent reasons to support the idea that referential processes — which are otherwise uncontroversially regarded as being semantically-mediated — are nonetheless sensitive to morphosyntactic agreement. Ginzburg (1992), who to our knowledge was the first to point this out, notes that in languages with certain types of gender systems (e.g., see examples that Culicover & Jackendoff (2005: 261, ex. 41c-e) provide for Icelandic, Russian, and Serbo-Croatian), the gender of a pronoun must agree not with the natural gender of the object being referred to, but instead the gender associated with the lexical item typically used to refer to that object. Revealingly, this constraint holds even if there is no antecedent, as in the case of exophora: Culicover & Jackendoff (2005) note that in Icelandic, for instance, reference to a book that is only situationally-evoked requires a feminine pronoun, only because the Icelandic word for *book* is feminine. Culicover & Jackendoff (2005) discuss such constraints in detail, and add similar cases from English presented by objects referred to with *plurale tantum* nouns (*scissors*, *pants*): again, even in cases of exophora, a plural pronoun is required even though the referents are notionally singular objects ([Pointing to scissors:] *Could you hand me <u>those</u>?*). Whereas this behavior imputes a role for morphosyntactic agreement in reference resolution (Culicover & Jackendoff (2005) appeal to a notion of INDIRECT LICENSING, which not only allows for orphan phrases to be licensed by a sentential antecedent, but also a lexical entry activated by such an antecedent, a lexical entry activated by the non-linguistic context, or a grammar rule that establishes a syntactic connection corresponding to the orphan's semantic role in the antecedent), presumably no linguist would take this as evidence that exophora involves reference to syntactic structure. In the end, it would seem that the lineage of syntactic analyses culminating in Chung (2013) has landed at a place quite similar to where non-transformational analyses started: positing constraints to deal with specific morphosyntactic facts having to do with case agreement and argument structure, but with no requirement for full syntactic identity.

With that said, both Chung (2013) and nontransformational analyses account for the apparent disconnect between argument sluices (as in (3a)–(3d)) and the adjunct sluices studied here: morphosyntactic constraints can lead to unacceptability in argument sluices, whereas no such constraints apply to adjunct sluices. Without wishing to take theoretical sides, we note that proponents of nontransformational approaches have provided independent motivation for their approach from the behavior of non-elliptical forms of discourse reference, and in this connection we see another potential advantage for these theories as well. Recall that, 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 seems to us to be more mysterious under a syntactic identity account than a referential one. 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 satisfying syntactic constraints that hold 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:[21] Whereas the penalty in question affected both elided and unelided variants in a way that was correlated within items, it was significantly exacerbated 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 (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).

When it comes to reference, on the other hand, it is well-known that the felicity of a speaker's choice to employ a particular referential form depends in part not only on the intended referent being available in the discourse context, but its degree of accessibility in the hearer's mental model of the discourse: the more linguistically reduced the referring expression is, the more accessible the referent needs to be (Ariel 1988; Gundel et al. 1993). Predictability of referent mention is one of the factors that affects accessibility, and the plausibility of the referent in turn affects predictability (Arnold 1998; Kehler & Rohde 2013). As such, a speaker's choice to employ ellipsis – the most linguistically reduced form of reference possible – to refer to a referent that has diminished accessibility would be expected to reduce sentence felicity, even if there were no other possible referents available in the discourse context. Whereas this observation obviously falls far short of providing an actual theory, it highlights another respect in which ellipsis behaves more like other referential processes, as opposed to a syntactic constraint satisfaction process.[22]

[21] See Merchant (2010) and Rudin (2019) for proposals along those lines.

[22] See Poppels (2022) for a more comprehensive discussion of the ways in which the behavior of ellipsis patterns with other forms of reference.

Finally, our results 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. Second, 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 suggests 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.

Supplementary files

All experimental items, data, data analysis scripts, and results can be accessed at https://github.com/tpoppels/poppels-kehler-sluicing-mismatch-paper.

Ethics and consent

All experiments reported in this paper, including the norming experiment described in the context of Experiment 2, were conducted under MIT IRB protocol #1605559077 – "Cognitive Foundations of Human Language Processing and Acquisition," which was approved by the Committee on the Use of Humans as Experimental Subjects (COUHES). The experiments were delivered in a web interface and provided participants with a written informed consent form. All participants gave their explicit consent through this form before the beginning of the experiment.

Funding information

We gratefully acknowledge support from National Science Foundation grant BCS-1456081 to Roger Levy.

Acknowledgments

We are grateful to three anonymous reviewers and to the editors for helpful comments on earlier drafts, and also participants of the 33rd Annual Human Sentence Processing Conference at UMASS Amherst (CUNY 2020) and the Experimental and Corpus-based Approaches to Ellipsis meeting hosted remotely at the University of Paris (ECBAE 2020).

Competing interests

The authors have no competing interests to declare.

References

Abels, Klaus. 2018. Movement & islands. In van Craenenbroeck, Jeroen & Temmerman, Tanja (eds.), *Handbook of ellipsis*, 389–424. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780198712398.013.17

Anand, Pranav & Hardt, Daniel & McCloskey, James. 2023. The domain of formal matching in sluicing. *Linguistic Inquiry (Early Access)*. DOI: https://doi.org/10.1162/ling_a_00495

AnderBois, Scott. 2010. Sluicing as anaphora to issues. In *Proceedings of Semantics and Linguistic Theory*, vol. 20. 451–470. DOI: https://doi.org/10.3765/salt.v20i0.2574

AnderBois, Scott. 2014. The semantics of sluicing: Beyond truth conditions. *Language* 90(4). 887–926. DOI: https://doi.org/10.1353/lan.2014.0110

Ariel, Mira. 1988. Referring and accessibility. *Journal of Linguistics* 24(1). 65–87. DOI: https://doi.org/10.1017/S0022226700011567

Arnold, Jennifier E. 1998. *Reference form and discourse patterns*. Stanford University dissertation.

Arregui, Ana & Clifton, Charles & Frazier, Lyn & Moulton, Keir. 2006. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language* 55(2). 232–246. DOI: https://doi.org/10.1016/j.jml.2006.02.005

Barker, Chris. 2013. Scopability and sluicing. *Linguistics and Philosophy* 36(3). 187–223. DOI: https://doi.org/10.1007/s10988-013-9137-1

Barr, Dale J. & Levy, Roger & Scheepers, Christoph & Tily, Harry J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. DOI: https://doi.org/10.1016/j.jml.2012.11.001

Barros, Matthew & Elliott, Patrick & Thoms, Gary. 2014. There is no island repair. Ms. Rutgers/UCL/University of Edinburgh.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1). 1–28. DOI: https://doi.org/10.18637/jss.v080.i01

Bürkner, Paul-Christian. 2018. Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 1–15. DOI: https://doi.org/10.32614/RJ-2018-017

Chung, Sandra. 2006. Sluicing and the lexicon: The point of no return. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, vol. 31. 73–91. Berkeley, CA. DOI: https://doi.org/10.3765/bls.v31i1.896

Chung, Sandra. 2013. Syntactic identity in sluicing: How much and why. *Linguistic Inquiry* 44(1). 1–44. DOI: https://doi.org/10.1162/LING_a_00118

Chung, Sandra & Ladusaw, William A. & McCloskey, James. 1995. Sluicing and logical form. *Natural Language Semantics* 3(3). 239–282. DOI: https://doi.org/10.1007/BF01248819

Chung, Sandra & Ladusaw, William & McCloskey, James. 2011. Sluicing(:) between structure and inference. *Representing language: Essays in honor of Judith Aissen,* 31–50.

Culicover, Peter W. & Jackendoff, Ray. 2005. *Simpler Syntax*. Oxford University Press on Demand. DOI: https://doi.org/10.1093/acprof:oso/9780199271092.001.0001

Dalrymple, Mary & Shieber, Stuart M. & Pereira, Fernando C. N. 1991. Ellipsis and higher-order unification. *Linguistics and Philosophy* 14(4). 399–452. DOI: https://doi.org/10.1007/BF00630923

Drummond, Alex. 2017. Ibex: Internet based experiments. Originally retrieved from https://spellout.net/ibexfarm.

Fiengo, Robert & May, Robert. 1994. *Indices and identity*. MIT press.

Frazier, Lyn. 2013. A Recycling Approach to Processing Ellipsis. In Cheng, Lisa Lai-Shen & Corver, Norbert (eds.), *Diagnosing Syntax*, 485–501. DOI: https://doi.org/10.1093/acprof:oso/9780199602490.003.0024

Frazier, Lyn & Duff, John. 2019. Repair or accommodation? Split antecedent ellipsis and the limits of repair. *Glossa: a Journal of General Linguistics* 4(1). DOI: https://doi.org/10.5334/gjgl.728

Gelman, Andrew & Hill, Jennifer & Yajima, Masanao. 2012. Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness* 5(2). 189–211. DOI: https://doi.org/10.1080/19345747.2011.618213

Ginzburg, Jonathan. 1992. *Questions, queries and facts: A semantics and pragmatics for interrogatives*: Stanford University dissertation.

Ginzburg, Jonathan & Sag, Ivan A. 2000. *Interrogative investigations*. Stanford: CSLI publications.

Gordon, Peter C. & Grosz, Barbara J. & Gilliom, Laura A. 1993. Pronouns, names, and the centering of attention in discourse. *Cognitive Science* 17(3). 311–347. DOI: https://doi.org/10.1207/s15516709cog1703_1

Gundel, Jeanette K. & Hedberg, Nancy & Zacharski, Ron. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307. DOI: https://doi.org/10.2307/416535

Hale, Kenneth & Keyser, Samuel Jay. 1993. On Argument Structure and the Lexical Expression of Syntactic Relations. In Hale, Kenneth (ed.), *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, 53–109. MIT Press.

Jäger, Gerhard. 2001. Indefinites and sluicing. A type logical approach. In van Rooy, Robert & Stokhof, Martin (eds.), *Proceedings of the 13th Amsterdam Colloquium*, 114–119.

Kehler, Andrew. 1993. The effect of establishing coherence in ellipsis and anaphora resolution. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, 62–69. DOI: https://doi.org/10.3115/981574.981583

Kehler, Andrew & Rohde, Hannah. 2013. A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics* 39(1–2). 1–37. DOI: https://doi.org/10.1515/tl-2013-0001

Kertz, Laura. 2010. *Ellipsis reconsidered*. San Diego, CA: University of California, San Diego dissertation.

Kertz, Laura. 2013. Verb phrase ellipsis: The view from information structure. *Language* 89(3). 390–428. DOI: https://doi.org/10.1353/lan.2013.0051

Kim, Christina S. & Kobele, Gregory M. & Runner, Jeffrey T. & Hale, John T. 2011. The Acceptability Cline in VP Ellipsis. *Syntax* 14(4). 318–354. DOI: https://doi.org/10.1111/j.1467-9612.2011.00160.x

Kim, Christina S. & Runner, Jeffrey T. 2018. The division of labor in explanations of verb phrase ellipsis. *Linguistics and Philosophy* 41(1). 41–85. DOI: https://doi.org/10.1007/s10988-017-9220-0

Kroll, Margaret. 2019. Polarity reversals under sluicing. *Semantics and Pragmatics* 12. 1–55. DOI: https://doi.org/10.3765/sp.12.18

Levin, Lori. 1982. Sluicing: A lexical interpretation procedure. In Bresnan, Joan (ed.), *The mental representation of grammatical relations*, 590–654. MIT Press.

Lipták, Anikó. 2015a. Identity in ellipsis: An introduction. *Lingua* 166. 155–171. DOI: https://doi.org/10.1016/j.lingua.2015.08.003

Lipták, Anikó. 2015b. Relative pronouns as sluicing remnants. *Approaches to Hungarian* 14. 187–207. DOI: https://doi.org/10.1075/atoh.14.08lip

Merchant, Jason. 2001. *The Syntax of Silence: Sluicing, islands, and the theory of ellipsis*. Oxford University Press on Demand. DOI: https://doi.org/10.1093/oso/9780199243730.001.0001

Merchant, Jason. 2005. Revisiting syntactic identity conditions. Talk presented at *the Workshop on Identity in Ellipsis*. UC Berkeley.

Merchant, Jason. 2008. An asymmetry in voice mismatches in VP-ellipsis and pseudogapping. *Linguistic Inquiry* 39(1). 169–179. DOI: https://doi.org/10.1162/ling.2008.39.1.169

Merchant, Jason. 2010. Three types of ellipsis. In Recanati, Francois & Stojanovic, Isidora & Villanueva, Neftalí (eds.), *Context-dependence, perspective and relativity*, 141–192. De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110227772.2.141

Merchant, Jason. 2013. Voice and ellipsis. *Linguistic Inquiry* 44(1). 77–108. DOI: https://doi.org/10.1162/LING_a_00120

Messick, Troy. 2012. Ellipsis and reconstruction in tough-infinitives. In *Proceedings of Generative Linguistics in the Old World in Asia IX*, 173–185.

Murphy, Andrew. 2020. Voice mismatches beyond passives: sluicing with active impersonal antecedents. *Linguistica Brunensia* 68(2). 63–85. DOI: https://doi.org/10.5817/LB2020-2-5

Nykiel, Joanna & Kim, Jong-Bok. 2022. Fragments and structural identity on a direct interpretation approach. *Journal of Linguistics* 58(1). 73–109. DOI: https://doi.org/10.1017/s0022226720000420

Poppels, Till. 2022. Explaining ellipsis without identity. *The Linguistic Review* 39(3). 341–400. DOI: https://doi.org/10.1515/tlr-2022-2091

Poppels, Till & Kehler, Andrew. 2019. Reconsidering asymmetries in voice-mismatched verb phrase ellipsis. *Glossa: A Journal of General Linguistics* 4. 1–22. DOI: https://doi.org/10.5334/gjgl.738

Poppels, Till & Kehler, Andrew. 2023. Ellipsis and the QUD: Sluicing with nominal antecedents. In Konietzko, Andreas & Winkler, Susanne (eds.), *Information structure and discourse in generative grammar: Mechanisms and processes*, 389–424. De Gruyter Mouton.

Potsdam, Eric. 2007. Malagasy sluicing and its consequences for the identity requirement on ellipsis. *Natural Language & Linguistic Theory* 25(3). 577–613. DOI: https://doi.org/10.1007/s11049-006-9015-4

R Core Team. 2021. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Ross, John. 1969. Guess who? In *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society,* 252–286.

Rudin, Deniz. 2019. Head-based syntactic identity in sluicing. *Linguistic Inquiry* 50(2). 253–283. DOI: https://doi.org/10.1162/ling_a_00308

Sag, Ivan A. 1976. *Deletion and logical form*: Massachusetts Institute of Technology dissertation.

Sag, Ivan A. & Nykiel, Joanna. 2011. Remarks on sluicing. In *Proceedings of the 18th international conference on head-driven phrase structure grammar*, 188–208. DOI: https://doi.org/10.21248/hpsg.2011.11

Tanaka, Hidekazu. 2011a. Syntactic identity and ellipsis. *The Linguistic Review* 28(1). 79–110. DOI: https://doi.org/10.1515/tlir.2011.003

Tanaka, Hidekazu. 2011b. Voice mismatch and syntactic identity. *Linguistic Inquiry* 42(3). 470–490. DOI: https://doi.org/10.1162/ling_a_00054

Thoms, Gary. 2015. Syntactic identity, parallelism and accommodated antecedents. *Lingua* 166. 172–198. DOI: https://doi.org/10.1016/j.lingua.2015.04.005

Yoshida, Masaya. 2010. "Antecedent-contained" sluicing. *Linguistic Inquiry,* 348–356. DOI: https://doi.org/10.1162/ling.2010.41.2.348