

Reconsidering asymmetries in voice-mismatched VP-ellipsis

Till Poppels
UC San Diego
9500 Gilman Dr, La Jolla, CA 92093
tpoppels@ucsd.edu

Andrew Kehler
UC San Diego
9500 Gilman Dr, La Jolla, CA 92093
akehler@ucsd.edu

Abstract Previous research on VP-ellipsis has revealed the existence of a Mismatch Asymmetry, whereby cases with passive voice ellipsis clauses and active antecedent clauses are less acceptable than cases with active ellipsis clauses and passive antecedents. According to the memory-based explanation offered by the Recycling Hypothesis (RH; Arregui et al., 2006), this effect arises because passive clauses are more prone to be misremembered as active than the other way around, and hence passive-active mismatches are more likely to create an ‘illusion of grammaticality’. This paper describes three experiments that explore the source of the asymmetry, with particular attention to the predictions of the RH account on previously unexamined cases. The findings are inconsistent with the predictions of memory-based explanations such as the RH, and instead point to the existence of a penalty against passive ellipsis clauses in subject focus environments, one that applies to both matched and mismatched cases of VP-ellipsis and in both anaphoric and cataphoric discourse configurations. A possible explanation for the penalty is offered as a subject for future work.

Keywords: VP-ellipsis; voice mismatch; sentence processing; memory retrieval; information structure

1 Introduction

When contextual conditions are right, natural languages allow speakers to omit linguistic material that expresses information that they nonetheless intend to communicate to their hearers. One way in which this manifests is through ELLIPSIS – the phenomenon by which linguistic material that the grammar would otherwise require is left missing. This leaves the hearer with the task of determining the meaning associated with this missing material, which is typically achieved by way of appeal to the surrounding linguistic (and in some cases, non-linguistic) context.

There are many types of ellipsis found in the world’s languages, with a diverse set of constraints governing their use. Our focus in this paper will be a form of ellipsis commonly found in English, VP-ellipsis, which is exemplified in (1):

- (1) John looked into the problem, and Bill did too.

Here, the stranded auxiliary *did* in the second clause (henceforth referred to as the ELLIPSIS clause) marks a vestigial verb phrase, a meaning for which must be identified to recover the proposition denoted by the sentence. In many cases of VP-ellipsis, the ability to identify the meaning is enabled by the occurrence of another linguistic expression (the ANTECEDENT), which in this case is located in the first clause.

Despite the considerable attention paid to VP-ellipsis in the literature, the conditions under which a representation of an utterance may serve as a suitable antecedent for interpreting it remain poorly understood. Indeed, there is even no broad consensus on what we will henceforth refer to as the **FUNDAMENTAL QUESTION** regarding VP-ellipsis: at what level(s) of language processing (e.g., syntax, semantics, information structure) do constraints on acceptable usage apply and interpretation mechanisms operate? On **SYNTACTIC ANALYSES** (Sag 1976; Van Craenenbroeck 2012; Merchant 2013; Thoms 2015; inter alia), for instance, the recovery of the elided VP meaning is dependent on there being a suitable syntactic VP to serve as an antecedent in the discourse. Such analyses typically (but not universally; see Merchant (2013)) predict that VP-ellipsis will only be acceptable (and indeed, grammatical) when a syntactically-matching VP is available in the immediate context. On **REFERENTIAL ANALYSES** (Schachter 1977; Webber 1978; Chao 1988; Hardt 1999; Lobeck 1999; Kehler 2000; inter alia), on the other hand, VP-ellipsis is treated as a null proform. Interpretation is thus predicted to be governed by the same types of processes used to resolve other types of referential expressions such as pronouns, which importantly are understood to be interpreted with respect to semantic representations of referents in the hearer's mental model of the discourse, rather than with respect to syntactic constituents.

A natural place to look to evaluate the competing predictions of the theories, therefore, is cases in which a salient VP meaning should be readily recoverable from the discourse context, but the antecedent is not in the required syntactic form. Syntactic analyses predict that such cases should be unacceptable, whereas referential analyses predict that they should be acceptable. The data, however, are notoriously unclear. Consider the following variants of (1):

- (2) a. #This problem was looked into by John, and Bob did too. [looked into the problem]
 b. This problem was to have been looked into, but obviously nobody did. [look into the problem] (Vincent Della Pietra, in conversation, cited in Kehler, 1993)

Example (2a) is a case of a voice mismatch: the ellipsis clause is in active voice, but the antecedent clause is in the passive. Therefore the VP necessary on the syntactic account (*looked into the problem*) is not present in the context, hence the example's unacceptability. The challenge for referential analyses is to explain this unacceptability given that a meaning corresponding to the missing VP should be made available by the meaning of the first sentence. Example (2b), on the other hand, is widely judged to be acceptable, despite it having the same passive-active mismatch that characterizes (2a). Such examples hence challenge syntactic analyses, since the required VP is not available in this case either.¹

The issues surrounding the status of examples that involve mismatch have inspired a considerable amount of experimental work that has sought to obtain more fine-grained measurements of acceptability, generally by utilizing acceptability rating tasks (Arregui et al. 2006; Kim, Kobele, et al. 2011; SanPietro, Xiang & Merchant 2012; Kertz 2013; Kim & Runner 2018; inter alia). This research has uncovered two significant patterns. The first is that cases involving mismatch such as (3b) are reliably judged to be less acceptable than paired variants in which the voice is matched as in (3a).

- (3) a. The judge read the report first, and then the lawyer did too. match
 b. #The report was first read by the judge, and then the lawyer did too. mismatch

¹ There are a variety of analyses on offer for reconciling these data (Kehler 2002; Kim, Kobele, et al. 2011; Grant, Clifton & Frazier 2012; Kertz 2013), the details of which will not concern us.

The second pattern, which is in fact the primary concern of our work, is the existence of differences in acceptability between different *types* of voice mismatches. Specifically, passive-voice VP-ellipsis with active antecedents, as in (4b), tends to be less acceptable than active voice VP-ellipsis with passive antecedents, as in (4a).

- (4) a. The report was first read by the judge, and then the lawyer did too. = (3b) [P → A]
 b. The judge read the report first, and then the confession was. [A → P]

This finding, which was first reported by Arregui et al. (2006: Experiment 5), has been replicated in several subsequent studies (Parker 2017; Kim & Runner 2018; Xiang & Klafka 2018).² We henceforth refer to this pattern as the MISMATCH ASYMMETRY. This finding gives rise to an immediate question: Does the Mismatch Asymmetry result from the linguistic properties of VP-ellipsis, and hence require a linguistic explanation, or does it reflect a fact about processing, external to the theory of VP-ellipsis itself? Pinning down the locus of the phenomenon is important for understanding both the linguistic constraints that govern the use of ellipsis as well as the mechanics of the interpretation process that comprehenders utilize to recover its meaning.

Several recent works have argued for a processing-based explanation, specifically based on the behavior of memory (Arregui et al. 2006; Parker 2017; Xiang & Klafka 2018). Here we focus specifically on the analysis of Arregui et al. (2006) (see also Frazier, 2013), who explain the mismatch asymmetry by appeal to a processing theory known as the Recycling Hypothesis (RH). The RH has two components: (i) a grammatical constraint on the use of VP-ellipsis, which requires syntactic identity between the elided material and its antecedent; and (ii) a processing theory to explain any residual variation in acceptability when the grammar predicts ungrammaticality. Given their notion of syntactic identity, voice-mismatched VP-ellipsis is categorically ruled out as ungrammatical, and as a result [P → A] and [A → P] mismatches are predicted to be equally unacceptable as far as the grammar is concerned. However, whenever the sentence processor is faced with a grammatical violation, it attempts to reanalyze past syntactic material and "recycle" it in a way that renders the input grammatical. In the case of ellipsis with non-identical antecedents, this Recycler is taken to reanalyze the existing antecedent and fashion an alternative antecedent that satisfies the identity constraint. The amount of work that the Recycler needs to carry out in order to repair an ellipsis is hypothesized to determine the relative level of acceptability of the passage in question, such that ungrammatical cases of ellipsis with non-identical antecedents may be perceived as relatively acceptable as long as an identical antecedent can be "recycled" from the existing one without a lot of effort.

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

² Note however that Kim, Koble, et al. (2011) failed to find a reliable effect; see their Experiment 1.

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

The same logic is then taken to apply to the hearer as well: previously heard passive clauses are more likely to be misremembered as having been in the active voice than previously heard active clauses are to be misremembered as having been in the passive (Mehler 1963). Since the grammar licenses ellipsis only when the elided material is syntactically identical to the antecedent, the processing of the second clause requires the retrieval of the first clause from memory in order to evaluate whether the two are identical. When the to-be-retrieved clause is passive, there is some chance that it will be misremembered as active, resulting in an "illusory" [A → A] match, which the authors refer to as an "illusion of grammaticality." This idea is illustrated in (5a), where greyed-out font indicates a relatively noisy memory trace of the passive antecedent clause.

- (5) a. The report was first read by the judge before the lawyer did too. illusory [A → A]
 b. The judge read the report first before the confession was too. [A → P]

Since active clauses are less prone to being misremembered as passive, mismatches as in (5b) are less likely to elicit such an illusion of grammaticality and are therefore, under the RH, predicted to receive a lower average acceptability rating. On this story, therefore, the effect is explained as a processing phenomenon, and hence requires no special accommodation within the theory of ellipsis itself.

We have now seen two sets of experimental data that potentially inform the fundamental question, and a representative answer from advocates of syntactic analyses – the RH – that seeks to explain the effects by way of a processing model that lies external to the grammar of VP-ellipsis. The question now is what sort of explanation could be offered under the posits of a referential theory. As argued by Kehler (2017; 2019), the existence of mismatch effects is not necessarily surprising on referential theories, since it is well-known from research on entity-level reference that the linguistic form of an antecedent expression can affect the relative level of accessibility of the entity it denotes with respect to the hearer's mental model of the discourse. Applying this idea to the case of reference to eventualities, we observe that when a syntactic match exists as in (6a), a representation of the meaning of the referent has already been computed by way of the compositional semantic analysis of the VP, per (6b).

- (6) a. John looked into the problem, and Bill did too.
 b. [[VP]: $\lambda x. look_into(x, problem)$
 c. The problem was looked into by John, and Bill did too.
 d. [[VP]: $\lambda x. look_into(John, x)$
 e. [[S]: $look_into(John, problem)$

That is to say, at the time that the ellipsis site is encountered, a representation of the referent will already be in the hearer's mental model of the discourse. This is not the case, however, when there is a syntactic mismatch as in (6c); here the compositionally-determined meaning of the VP, shown in (6d), is not the required one. Obtaining the necessary meaning (6b) will require an additional computation, e.g., the recovery of a lambda abstract from a representation of the meaning of the entire clause (6e).³ So the idea that a modicum of additional discourse-level processing might

³ For a procedure that resolves all VP-ellipses by way of such a calculation, see Dalrymple, Shieber & Pereira (1991).

be required to fashion a representation of the referent in the case of syntactic mismatches, under the presumption that VP-ellipsis presupposes that the representation is already available, could potentially explain their reduced acceptability.

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

This reasoning only goes through, of course, if the RH's analysis of the Mismatch Asymmetry is the correct one. The goal of this paper is therefore to explore the source of the asymmetry, with particular attention to the predictions of the RH account on previously unexamined cases. Experiment 1 aims to replicate the Mismatch Asymmetry using stimuli adapted from the original [Arregui et al. \(2006\)](#) study, but to do so using a fuller paradigm that contains voice-matched control items that were not included in the original experiment. Whereas the results succeed in replicating the key finding of Arregui et al.'s study, they also reveal a penalty for passive ellipsis clauses even when the ellipsis clause is syntactically matched with the antecedent clause, suggesting the existence of a more general passive penalty for ellipsis clauses. Experiment 2 considers the question of whether the passive penalty might be independent of ellipsis *per se*, in part by examining the acceptability of unelided variants of the stimuli used in Experiment 1. Whereas the effect found in Experiment 1 was replicated for the elided versions, no such effect was found for the unelided variants, indicating that the passive penalty is specific to ellipsis clauses. Experiment 3 then provides the critical test of the RH analysis by examining cases that feature cataphoric VP-ellipsis. In such cases, the RH and the passive penalty hypothesis make opposing predictions: the passive penalty hypothesis predicts that mismatches with passive ellipsis clauses should remain worse than those with active ellipsis clauses, whereas the RH predicts that the judgments should reverse, since it is now the structure of the ellipsis clause, by virtue of occurring first, that is subject to misremembering. The results support the existence of a more general passive penalty for ellipsis clauses as opposed to a memory-based explanation. As an ensemble, the results therefore suggest that neither memory-based explanations such as the RH nor any ellipsis-independent explanation is capable of accounting for the Mismatch Asymmetry. We conclude by discussing some ramifications of our results for the debate between syntactic and referential theories of VP-ellipsis, as well as a possible source of the effect.

2 Experiment 1

As described above, [Arregui et al. \(2006\)](#) found a Mismatch Asymmetry whereby [A → P] mismatches in VP-ellipsis were subject to a greater acceptability penalty than were [P → A] mismatches. However, no matched controls were included, which are necessary to confirm that the effects are specific to mismatched cases.⁴ Because the RH explains the mismatch asymmetry as

⁴ Indeed, other recent work suggests that the effects might not be, although the evidence isn't unequivocal. For instance, [Kim, Koble, et al. \(2011\)](#) found that cases of [P → P] matched VP-ellipsis were rated as less acceptable than [A

being a by-product of the Recycling process, and it explicitly bars the Recycler from being recruited unless there is a grammatical violation, one expects that the difference between active and passive clauses in terms of memory retrieval will have no effect on voice-*matched* VP-ellipsis. The purpose of Experiment 1 is thus twofold: to replicate the Mismatch Asymmetry that Arregui et al. found, and to include voice-matched controls to further examine the predictions of the analysis.

2.1 Methods

2.1.1 Stimuli

Twenty-four experimental items followed a 2x2 design, crossing two independent factors: whether the ellipsis clause and the antecedent MATCHED or MISMATCHED in voice, and whether the ellipsis clause was ACTIVE or PASSIVE, as illustrated in (7)–(8).

- | | | | |
|-----|----|--|----------|
| (7) | a. | The judge read the report first, and then the lawyer did too. | [A -> A] |
| | b. | The report was first read by the judge, and then the confession was too. | [P -> P] |
| | c. | The report was first read by the judge, and then the lawyer did too. | [P -> A] |
| | d. | The judge read the report first, and then the confession was too. | [A -> P] |
| (8) | a. | The customer praised the dessert after the critic did already. | [A -> A] |
| | b. | The dessert was praised by the customer after the appetizer was already. | [P -> P] |
| | c. | The dessert was praised by the customer after the critic did already. | [P -> A] |
| | d. | The customer praised the dessert after the appetizer was already. | [A -> P] |

The mismatch variants were identical to the stimuli used in Arregui et al. (2006: Experiment 5), setting aside the correction of a small number of typos and a few changes so as to ensure that all clauses were plausible and identical across item variants. The voice-matched variants were constructed by holding the ellipsis clauses of the mismatched variants constant and exchanging the antecedent clauses, leaving everything else unchanged. As with Arregui et al.'s stimuli, in half of the stimuli the antecedent and ellipsis clauses were conjoined with *and (then)...too* as in (7), and in half they were in a subordinating configuration using the connective *after* (8). The items were supplemented with 48 filler items exemplified in (9): 24 acceptable fillers and 24 unacceptable ones, half of each involving ellipsis and half not.

- | | | | |
|-----|----|---|-------------------------------------|
| (9) | a. | The thief was arrested and his brother was as well. | Acceptable, elliptical filler |
| | b. | A proof that God exists doesn't. ⁵ | Unacceptable, elliptical filler |
| | c. | I can't hear the announcement but I don't care. | Acceptable, non-elliptical filler |
| | d. | What did you meet a janitor that hates? | Unacceptable, non-elliptical filler |

-> A] matched cases, but do not discuss the effect further. Unpublished studies by Parker (2017) and Xiang & Klafka (2018) report similar effects for their stimuli. On the other hand, Kim & Runner (2018) report a significant interaction between mismatch and antecedent voice, suggesting that voice-matched [A -> A] and [P -> P] stimuli did not differ in acceptability, although the data analyzed included non-elliptical variants as well. Hence we seek to investigate the question ourselves, using variants of Arregui et al.'s own stimuli.

⁵ This particular item is due to Sag (1976), who points out that it appears to be ungrammatical.

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

2.1.2 Procedure

We recruited 30 participants via Amazon.com’s Mechanical Turk, one of whom reported being a non-native English speaker and was excluded from all analyses. In a within-item and within-participant design, each participant was presented with exactly one variant of each of the 24 experimental items, which were presented in a random order and interspersed with the 48 filler items exemplified in (9). The materials were presented using the Ibex software for conducting psycholinguistic experiments online⁶ and participants were instructed to rate each item in terms of its acceptability on a scale from 1-5, with a 5 rating meaning that “the sentence is perfectly acceptable in English and that you can imagine yourself or other native speakers saying it.”

2.2 Predictions

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

2.3 Results

The results are summarized in Figure 1. We fit a linear mixed-effects regression model to the raw scores obtained in the acceptability judgment task using the `lme4` package (Bates et al. 2015) for R (R Development Core Team 2009), with MATCH/MISMATCH and VOICE of the ellipsis clause as fixed effects alongside the interaction between the two, and random intercepts and slopes for participants and items (Barr et al. 2013).⁷ The data reveal a significant mismatch penalty whereby examples with syntactically mismatched antecedent and ellipsis clauses were judged to be worse than examples in which the clauses were syntactically matched ($\beta = -0.41$, $SE = 0.07$, $t = -5.49$, $p < 0.001$). Further, [A → P] mismatches were less acceptable than [P → A] mismatches, which is reflected in a main effect whereby items with passive ellipsis clauses were significantly degraded compared to those with active ellipsis clauses ($\beta = -0.24$, $SE = 0.06$, $t = -3.86$, $p < 0.001$). These two main effects were independent of each other: we found no evidence for an interaction between the two ($\beta = -0.03$, $SE = 0.05$, $t = -0.61$, $p = 0.55$).

⁶ <https://github.com/addrummond/ibex>

⁷ The complete formula was: `response ~ mismatch + voice.ellipsis + mismatch:voice.ellipsis + (1 + mismatch*voice.ellipsis | subject) + (1 + mismatch*voice.ellipsis | item)`.

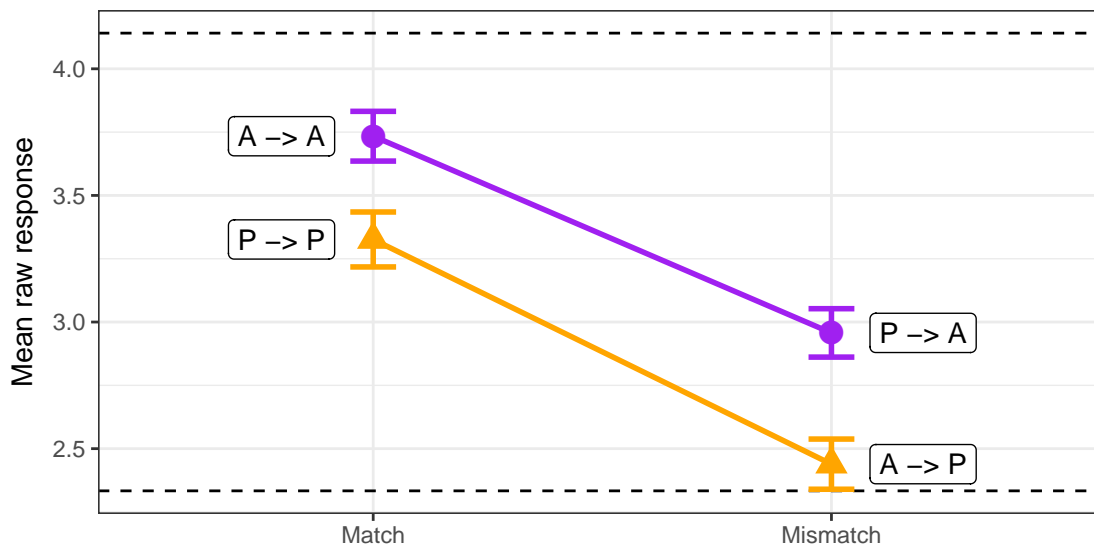


Figure 1: Results from Experiment 1. Dashed lines indicate mean acceptability of (un)acceptable elliptical fillers. Errorbars show Standard Errors.

2.4 Discussion

The goal of Experiment 1 was to evaluate the three predictions outlined in Section 2.2. Consistent with previous findings and the predictions of the RH, the results confirmed the first prediction, whereby syntactically mismatched cases of VP-ellipsis were reliably judged to be less acceptable than syntactically matched cases. The second prediction was also borne out: as expected, [A → P] mismatches were less acceptable than [P → A] mismatches, replicating the Mismatch Asymmetry effect identified by Arregui et al. The results failed to confirm the third prediction of the RH, however, according to which there should be no analogous effect in syntactically matched cases. Instead, a parallel effect was in fact found, whereby [P → P] matches were rated as less acceptable than [A → A] matches, with no interaction.

The results therefore cast doubt on the RH's memory-based explanation of the Mismatch Asymmetry. Instead, they suggest that the Mismatch Asymmetry is driven by the existence of a more general, and hence mismatch-independent, penalty for passive ellipsis clauses, one that affects acceptability regardless of whether the antecedent and ellipsis clause differ in voice. Importantly, the data do not support the existence of a penalty against passive clauses more generally, as that would predict no difference between the [P → A] and [A → P] mismatches, as well as a greater degree of degradation of [P → P] matches as compared to the other conditions. Instead, the data is explained best by an additive combination of a mismatch penalty and a passive ellipsis clause penalty.

3 Experiment 2

The results from Experiment 1 appear to be problematic for the RH, since the hypothesis offers no explanation for why a parallel penalty for passive ellipsis clauses would be witnessed in the matched condition. The results instead support the existence of a more general passive ellipsis

clause penalty (henceforth, PECP) – whatever the underlying explanation for it might be – and that the mismatch asymmetry is merely a by-product of that effect.

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

3.1 Methods

3.1.1 Stimuli

In addition to the stimuli used in Experiment 1 exemplified in (7) and (8), a no-ellipsis condition was added with variants in which the elided VP was made overt, as in (10). Following Kim & Runner (2018), the overt VP was reduced linguistically as much as possible, e.g. by pronominalizing NPs wherever it was felicitous to do so, in order to mitigate a potential independent penalty associated with producing overt material that could have been elided.

- (10) a. The judge read the report first, and then the lawyer read it too. [A – A]
 b. The report was first read by the judge, and then the confession was read too. [P – P]
 c. The report was first read by the judge, and then the lawyer read it too. [P – A]
 d. The judge read the report first, and then the confession was read too. [A – P]

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

3.1.2 Procedure

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

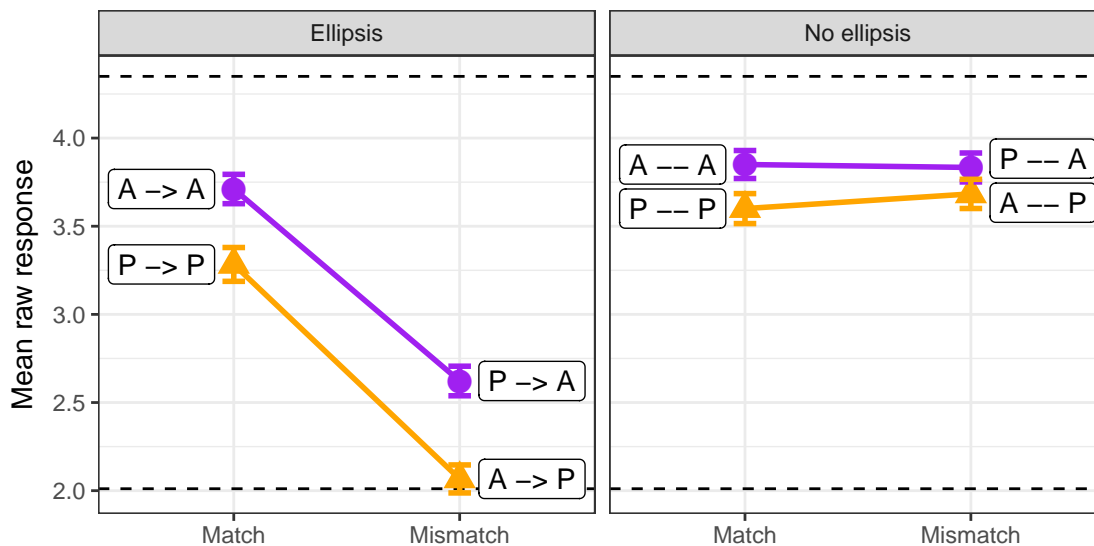


Figure 2: Results from Experiment 2. Errorbars reflect Standard Errors, dashed lines indicate mean ratings of (un)acceptable elliptical fillers.

3.2 Results

We conducted two mixed-effects regression analyses that were analogous to the one for Experiment 1, each with MISMATCH and VOICE of the ellipsis clause as fixed effects along with the interaction between the two, and the maximal random effect structure for items and subjects. The results are summarized in Figure 2. First, considering only data in the ellipsis condition (left panel), we fully replicated the results from Experiment 1, revealing significant mismatch ($\beta = -0.57$, $SE = 0.07$, $t = -7.72$, $p < 0.001$) and passive penalties ($\beta = -0.25$, $SE = 0.062$, $t = -4.04$, $p < 0.001$), and no evidence for an interaction between the two ($\beta = -0.03$, $SE = 0.04$, $t = -4.04$, $p = 0.47$). The no-ellipsis condition (right panel), however, revealed no evidence for a mismatch penalty ($\beta = 0.02$, $SE = 0.04$, $t = 0.043$, $p = 0.67$), only a small, statistically marginal passive penalty ($\beta = -0.1$, $SE = 0.05$, $t = -1.9$, $p = 0.068$), and no interaction ($\beta = 0.03$, $SE = 0.05$, $t = 0.68$, $p = 0.5$).

3.3 Discussion

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

4 Experiment 3

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

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

In order to more definitively compare the two explanations of the Mismatch Asymmetry, however, a type of example is needed for which the hypotheses make both crisp and opposing predictions. Fortunately, there is such a case. As is well-known, VP-ellipsis is acceptable when used cataphorically in subordinate discourse configurations, as in (11):

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

Example (11) is acceptable even though the ellipsis clause *he wants to* precedes the catecedent clause *the judge will read the report*. This referential pattern mirrors that of the pronoun in (11), whereby reference with *he* is successful despite preceding its catecedent *the judge*.

The PECP hypothesis and RH make opposite predictions for such cases. On the one hand, the predictions of the PECP hypothesis are as before: mismatches that contain a passive ellipsis clause (and hence an active catecedent clause) should be judged as less acceptable than those involving an active ellipsis clause and a passive catecedent. That is, the ordering of the clauses shouldn't matter. The RH, on the other hand, makes the opposite prediction: mismatches that contain a passive ellipsis clause and active catecedent clause should be judged as *more* acceptable than those that contain an active ellipsis clause and a passive catecedent. This prediction results from the fact that, by virtue of being the initial clause, it is the *ellipsis* clause that is subject to misremembering. That is, upon encountering a cataphoric ellipsis site, the processor will anticipate, and ultimately identify, the occurrence of the catecedent. When the processor attempts to establish identity between the catecedent and the (invisible) structure at the ellipsis site, it is the ellipsis clause that has to be retrieved from memory. As illustrated in (12), since [P ← A] mismatches require the retrieval of a passive clause from memory, they are more likely to elicit an illusion of grammaticality than [A ← P] mismatches, which involve the retrieval of an active clause.

(12) a. Before the lawyer did, the report was first read by the judge. [A ← P]

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

The purpose of Experiment 3 is to evaluate these competing predictions.⁸

4.1 Methods

4.1.1 Stimuli

Cataphoric variants of the stimuli used in Experiment 1 were constructed. Recall that half the stimuli from Experiment 1 employed *after* to connect the clauses, as in (8). Since *after* is a subordinating conjunction, the variants could be constructed simply by reversing the order of the clauses, as in (13).

- (13) a. After the critic did already, the customer praised the dessert. [A <- A]
 b. After the appetizer was already, the dessert was praised by the customer. [P <- P]
 c. After the critic did already, the dessert was praised by the customer. [A <- P]
 d. After the appetizer was already, the customer praised the dessert. [P <- A]

On the other hand, the other half of the stimuli used in Experiment 1 employed the coordinating conjunction *and*, and hence the second clause could not be fronted to form a subordinate structure. We therefore adapted the examples to employ the subordinate connective *before*, as in (14).

- (14) a. Before the lawyer did, the judge read the report first. [A <- A]
 b. Before the confession was, the report was first read by the judge. [P <- P]
 c. Before the lawyer did, the report was first read by the judge. [A <- P]
 d. Before the confession was, the judge read the report first. [P <- A]

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

4.1.2 Procedure

As in Experiment 1, we recruited 30 participants from Amazon.com's Mechanical Turk and presented with 24 experimental stimuli alongside 48 filler items in an acceptability judgment task via

⁸ A reviewer questions whether the predictions of the RH for cataphora are as straightforward as our characterization would suggest, noting that it is possible that cataphoric and non-cataphoric cases are processed in different ways. Specifically, the reviewer suggests that the forward-looking dependency created by cataphoric VP-ellipsis might lead to stronger maintenance of the initial clause in active memory when the matrix clause is processed. An increase in the memory trace of this sort would in turn predict that mismatches that contain a passive ellipsis clause and an active catacedent should be as (un)acceptable as those with an active ellipsis clause and a passive catacedent, since cataphoric passive ellipsis clauses would not be subject to the misremembering effect.

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

Ibex. Two participants reported being non-native English speakers and were therefore excluded from all analyses. The details of the design and task followed those given in Section 2.1.2.

4.2 Predictions

Two predictions carry over from Experiment 1. First, we expect an effect of mismatch, whereby [A <- P] and [P <- A] cases are judged to be less acceptable than [A <- A] and [P <- P] cases. Second, the PECP hypothesis, but not the RH, predicts that [P <- P] examples will be judged as less acceptable than the [A <- A] examples. On the hypothesis that the PECP is an independent factor that co-exists with the mismatch penalty, the PECP hypothesis does not predict an interaction.

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

4.3 Results

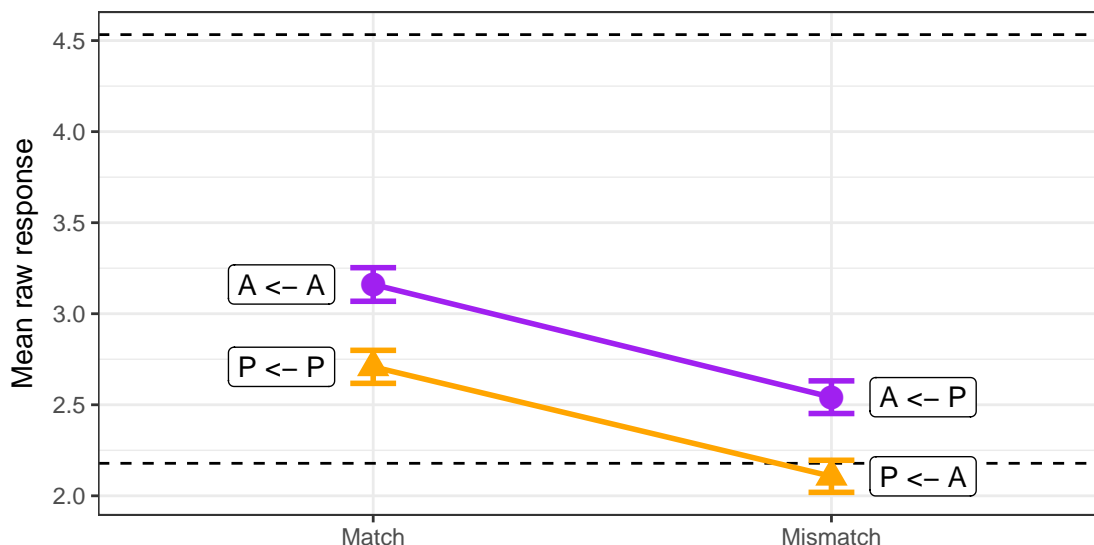


Figure 3: Results from Experiment 3. Errorbars indicate Standard Errors, and dashed lines show mean ratings of (un)acceptable elliptical fillers.

As in Experiments 1 and 2, we analyzed raw acceptability in a linear mixed-effects regression with MISMATCH and VOICE of the ellipsis clause and the interaction between the two as fixed effects and the maximal random effects structure permitted by the design. The results from Experiment 3 are summarized in Figure 3. As in Experiment 1, there were two significant main effects: a mismatch penalty ($\beta = -0.3$, $SE = 0.14$, $t = -5.92$, $p < 0.001$), and a penalty for passive ellipsis clauses ($\beta = -0.23$, $SE = 0.05$, $t = -4.55$, $p < 0.001$). There was no evidence that the

passive penalty differed across matched/mismatched conditions ($\beta = 0.007$, $SE = 0.04$, $t = 0.16$, $p = 0.87$).

4.4 Discussion

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

5 General discussion

We set out to explore the source of the Mismatch Asymmetry, in part by evaluating the predictions of the RH account on previously unexamined cases. Experiment 1 sought to replicate the asymmetry using stimuli adapted from the original [Arregui et al. \(2006\)](#) study, but using a fuller paradigm that contained voice-matched control items. Whereas the results replicated the key finding of Arregui et al.'s study, they also revealed a penalty for passive ellipsis clauses even when the ellipsis clause is syntactically matched with the antecedent clause. This finding instead suggested the existence of a passive ellipsis clause penalty, or PECP. Experiment 2 then asked whether the passive penalty might be attributable to a more general, ellipsis-independent condition by also examining unelided variants. Whereas the effect found in Experiment 1 was replicated for the elided versions, no such effect was found for the unelided variants, indicating that the passive penalty is specific to ellipsis clauses. Experiment 3 then provided a critical test of the RH analysis by examining cases that feature cataphoric VP-ellipsis. The results revealed that, like the previous two experiments, mismatches with passive ellipsis clauses are rated as worse than those with active ellipsis clauses. This result runs counter to the RH, as it predicts the opposite effect. Together, therefore, the experiments point to the existence of a PECP that applies across the board to both matched and mismatched cases of VP-ellipsis, in both anaphoric and cataphoric discourse configurations.

In addition to offering a refutation of the RH's explanation of the Mismatch Asymmetry, the results of our study call into question the relevance of the asymmetry to the fundamental question set out in the introduction, specifically regarding at what level(s) of language processing constraints on acceptable usage apply and interpretation mechanisms operate. For one, recall that if the predictions of the RH were confirmed, it would have potentially provided strong support for syntactic analyses, since the misremembering phenomenon to which the RH appeals applies specifically to syntactic representations. That is, no similar explanatory path would appear to be available to proponents of referential theories. Our findings cast significant doubt on the efficacy of the analysis, however, with the result being that this line of argumentation in favor of syntactic accounts is ren-

dered moot. To be clear, the results presented here do not argue *against* syntactic analyses either. Instead, the Mismatch Asymmetry remains a mystery on both syntactic and referential analyses.

A second finding of note is the degraded acceptability of syntactically-matched passive voice ellipses as compared to matched active cases, as found in previous studies and Experiment 1. This result is also surprising for both types of account. It is mysterious on syntactic analyses, since for both [A → A] and [P → P] ellipses, there exists a syntactically-matching, and hence perfectly suitable, VP available in the syntactic representation of the antecedent clause. As such, there is no constraint violation involved, and hence no need for a recovery mechanism such as the RH. The finding is likewise mysterious for referential accounts, for analogous reasons. That is, in each scenario a suitable representation of the referent has been computed as part of the compositional semantic analysis of the antecedent clause, and hence should be readily available as a referent for a subsequent VP-ellipsis. In neither case are any additional inferential steps needed to fashion an appropriate representation of the referent, as we saw is necessary in cases of syntactic mismatch. The results therefore suggest that there is a penalty for passive ellipsis clauses, one that demands an explanation regardless of which type of account of VP-ellipsis interpretation one adopts.

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

On the one hand, it is well-known that active voice and passive voice constructions differ with respect to their information structural properties. Whereas the active voice construction in English is relatively unmarked with respect to information structure (with a relatively weak tendency for subjects to be construed as topics), one of the primary functions of the passive is to mark its subject as being topical (Shibatani 1985; Givón 1990; Rohde & Kehler 2014; *inter alia*). As such, whereas the meaning of any constituent could potentially be topical in (15a), there is a much stronger presumption that the report is topical in (15b).

- (15) a. The judge read the report.
b. The report was read by the judge.

Otherwise, it is unclear what a speaker's motivation would be to choose the passive over the unmarked active. Another way to cast the observation is in terms of Question-Under-Discussion (QUD) models of discourse coherence (Roberts 2012), according to which topical elements of an utterance are those meanings that are provided by the operative QUD. In this regard, we note that sentence (15a) could serve as a felicitous answer to a variety of implicit questions, e.g., *What happened?* or *What did the judge do?*. Sentence (15b), on the other hand, comes across as a better answer to the question *What happened to the report?*, in which the report is part of the topic. It would be a less natural answer to the question *What did the judge do?*, for instance.

On the other hand, recall that according to referential theories, VP-ellipsis is predicted to behave like other proforms, such as entity-referring personal pronouns.⁹ According to some theories

⁹ Although not unequivocal, there is a robust set of evidence that reveals a close patterning between VP-ellipsis and entity-level pronouns, the latter of which is widely agreed to be a semantically, rather than syntactically, mediated phenomenon. For instance, the very possibility of cataphoric reference, as exemplified in (13), is one such defin-

of pronoun usage (Gundel, Hedberg & Zacharski 1993; Grosz, Joshi & Weinstein 1995; Rohde & Kehler 2014; inter alia), pronouns serve an information structural function as well, specifically to indicate a continuation of an entity-level topic. For this reason, whereas the pronoun *He* in passage (16) is to some degree ambiguous between John and Bill, the pronoun *He* in (17), where the choice to use the passive has placed Bill in a strongly topical position, is more likely to be understood to refer to Bill than to John.

(16) John reprimanded Bill. He was upset.

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

On this logic, if VP-ellipsis is a proform, we expect it to likewise carry a presupposition that its meaning is topical. And this is indeed the case in the stimuli used here and in the previous studies surveyed. Viewing the question again through the lens of QUD analyses, passages like (18) cohere by virtue of their clauses each providing a partial answer to a common QUD, in this case, *What was read by the judge?*

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

The meaning of the elided VP is presupposed by the operative QUD, it is therefore topical in discourses such as (18).

As a result, conflicting demands occur in a passivized ellipsis clause such as that in (18): the VP-ellipsis requires that the VP meaning (*was read by the judge*) be topical, whereas the speaker has used a construction that indicates that the remnant subject NP (*the confession*) is topical. Clearly one cannot have it both ways, for focus must be present somewhere in the clause. In such a situation, the speaker therefore has at least two other options, both of which are preferable:

ing characteristic. Cataphoric VP-ellipsis is allowable when it is embedded as in sentence (ia), just like cataphoric reference with the pronoun *he* is in sentence (ib).

- (i) a. If the judge wants to, he will read the report.
- b. If he wants to be enlightened, the judge will read the report.
- c. #The judge wants to, and so he will read the report.
- d. #He wants to be enlightened, and so the judge will read the report.

On the other hand, cataphora is not allowable when the ellipsis is not embedded as in sentence (ic), as is the case for pronominal reference in sentence (id). This alignment between pronouns and VP-ellipsis is predicted straightforwardly on a referential theory.

Other litmus tests for referential processes include the ability to get non-local antecedents (consider the fact that both the VP-ellipsis and the pronoun in (ii) find their antecedents two clauses back), and that of split antecedents (consider the fact that both the VP-ellipsis and pronoun *them* get their referents from multiple, distinct antecedents in (iii)).

- (ii) The thought came back, the one nagging at him these past four days. He tried to *stifle it*. But the words were forming. He knew he *couldn't*. (Hardt 1990)
- (iii) Mary wants to go to Spain and Fred wants to go to Peru, but because of limited resources, only one of them can. (Webber 1978)

These and other types of behaviors (including exophora; see Miller & Pullum (2014) for a compelling treatment) argue for the discursual nature of pronominal reference, and hence argue for the discursual nature of VP-ellipsis as well. Notably, these behaviors are not shared by a variety of other, more clearly local and syntactically-governed forms of ellipsis, such as stripping and gapping. See Kehler (2019) for further discussion.

use the active voice construction, for which the elided VP meaning can felicitously serve as the topic, or use the passive without employing ellipsis, so that the surface subject can serve as the topic without conflicting constraints. This hypothesis thus gives us an explanation for why we see a PECP in both matched and mismatched conditions, but only when VP-ellipsis occurs.

As noted earlier, this hypothesis is only speculative, and it is not among our goals to offer a vigorous defense of it here. However, we do point out that it does make an immediate prediction: that there should be no penalty for passive ellipsis clauses in which the subject and VP meaning can be both construed as being part of the topic. There is in fact evidence to support this prediction. Kertz (2013) conducted an experiment to test her hypothesis that the penalty for syntactic mismatch would vary according to information structural properties associated with ellipsis clauses: in particular, that ellipsis clauses that display AUXILIARY FOCUS will be more resilient to mismatch than those that display SUBJECT FOCUS. Her Experiment 3 employed stimuli of the sort shown in (19a)–(19d):¹⁰

- (19) a. The technicians didn't install the line as quickly as the engineers did.
[subject focus, match]
- b. The line wasn't installed by the technicians as quickly as it could have been.
[auxiliary focus, match]
- c. The line wasn't installed by the technicians as quickly as the engineers did.
[subject focus, mismatch]
- d. The technicians didn't install the line as quickly as it could have been.
[auxiliary focus, mismatch]

Kertz compared the relative acceptability of cases in which accent falls on the subject of the ellipsis clause, as in (19a) and (19c), with cases in which accent falls on the auxiliary, as in (19b) and (19d). The results revealed a reliable interaction between mismatch and focus, whereby simple focus ellipses were rated as significantly more acceptable in the mismatch condition, but not in the match condition.

Two of Kertz's findings provide preliminary support for our hypothesis. First, in contrast to the results of our Experiment 1, there was no passive ellipsis clause penalty witnessed in the matched condition: [P → P] ellipses such as (19b) were not reliably rated as less acceptable than [A → A] ellipses such as (19a). This is explained by the fact that her [P → P] stimuli featured ellipsis clauses in which focus resided only on the auxiliary. As such, both the passivized subject and the VP meaning are topical, and no PECP resulted. Second, whereas her results revealed a mismatch asymmetry, it went in the opposite direction as the one found here and in other previous work: [A → P] mismatches such as (19d) were rated as *more* acceptable than [P → A] mismatches such as (19c). Again, this is consistent with the hypothesis, since her [A → P] mismatches, unlike her [P → A] mismatches, were cases of auxiliary focus, and hence the other elements of the sentence, including the meanings of both the subject NP and the elided VP, were topical. This suggests that there was no PECP in effect to bring down the ratings of the [A → P] mismatches.

Therefore, as an ensemble, the foregoing evidence suggests that there is a PECP, but one that applies only in ellipsis clauses that bear subject focus. These are just the cases in which the need for the meanings of both the subject and the elided NP to be topical are in conflict: the elision of the VP in turn requires that focus falls on the subject, whereas one of the central functions of the

¹⁰ There were also two non-ellipsis variants included in each stimulus set, which we omit here for simplicity.

passive is to mark its subject as topical. This conflict does not exist in auxiliary focus constructions, and hence we find no evidence of a penalty.¹¹

A reviewer rightfully asks whether the PECP is unique to VP-ellipsis, or if it extends to other types of ellipsis more generally. To gain insight into this question, we carried out a pilot experiment to investigate the potential existence of a PECP in two other forms of ellipsis, specifically gapping and sluicing. 24 items were derived from the items used in Experiment 1, with 12 containing gapping as in (20) and 12 containing sluicing as in (21).

- (20) a. Mary scolded Wilma, and Susan, Nancy.
 b. Wilma was scolded by Mary, and Nancy, by Susan.
 c. Mary scolded Wilma, and Susan scolded Nancy.
 d. Wilma was scolded by Mary, and Nancy was scolded by Susan.
- (21) a. Someone read the report, but I don't know who.
 b. The report was read by someone, but I don't know by whom.
 c. Someone read the report, but I don't know who read it.
 d. The report was read by someone, but I don't know by whom it was read.

Each item followed a 2x2 design that crossed VOICE (active vs. passive) with ELLIPSIS (ellipsis vs. no ellipsis).¹²

26 native speakers of English, recruited via Amazon.com's Mechanical Turk, participated in the pilot.

The results are shown in Figure 4. As for the previously reported experiments, we used the `lme4` R package to test for the presence of main effects of voice and ellipsis, considering gapping and sluicing items separately. Since there were only 12 items for each ellipsis type instead of 24, the pilot study has substantially lower statistical power than the main experiments. Consequently, our models with maximal random effects failed to converge. In keeping with the recommendations

¹¹ It is therefore perhaps unsurprising that attested cases of [A -> P] mismatches cited in the literature, such as (i)–(ii) from Kehler (2002), are characterized by auxiliary focus rather than subject focus:

- (i) Actually I have implemented it [= a computer system] with a manager, but it doesn't have to be. [implemented with a manager] (Steven Ketchpel, in conversation)
- (ii) Just to set the record straight, Steve asked me to send the set by courier through my company insured, and it was. [sent by courier through my company insured] (posting on the Internet)

¹² The design did not include cases that involve syntactic mismatch, in part because it is not clear how to do so for sluicing. Specifically, whereas it is possible to construct cases that unambiguously involve an active-passive mismatch as in (i),

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

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

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

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

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

in Barr et al. (2013), we fit separate models for each hypothesis test, each time removing all random slopes except for the one corresponding to the hypothesis under investigation. This way, we achieved convergence in all models except for the model testing for a main effect of voice, for which we removed all random slopes. Across all random-effect specifications we tried, none of the models varied meaningfully.

For sluicing (right side of Figure 4), whereas there was a numerical difference between active and passive item variants (vertical distance between lines in the graph), the statistical analysis revealed that it was not significant ($\beta = -0.2$, $SE = 0.15$, $t = -1.33$, $p = 0.19$). There was, however, a significant difference between elliptical and non-elliptical item variants in that items involving sluicing significantly improved in acceptability compared to their non-elliptical counterparts ($\beta = 0.19$, $SE = 0.05$, $t = 3.5$, $p = 0.002$). The interaction between ellipsis and voice was not significant ($\beta = 0.04$, $SE = 0.05$, $t = 0.69$, $p = 0.49$).

For gapping (left side of Figure 4), ellipsis had a large, negative effect on acceptability ($\beta = -0.65$, $SE = 0.11$, $t = -5.65$, $p < 0.001$). There was no main effect of voice ($\beta = 0.2$, $SE = 0.15$, $t = 1.33$, $p = 0.19$). The interaction between voice and ellipsis reached marginal significance ($\beta = 0.2$, $SE = 0.1$, $t = 1.88$, $p = 0.07$), whereby items involving passive gapping tended to be more acceptable than those involving active gapping, but with no such difference between their respective non-elliptical counterparts.

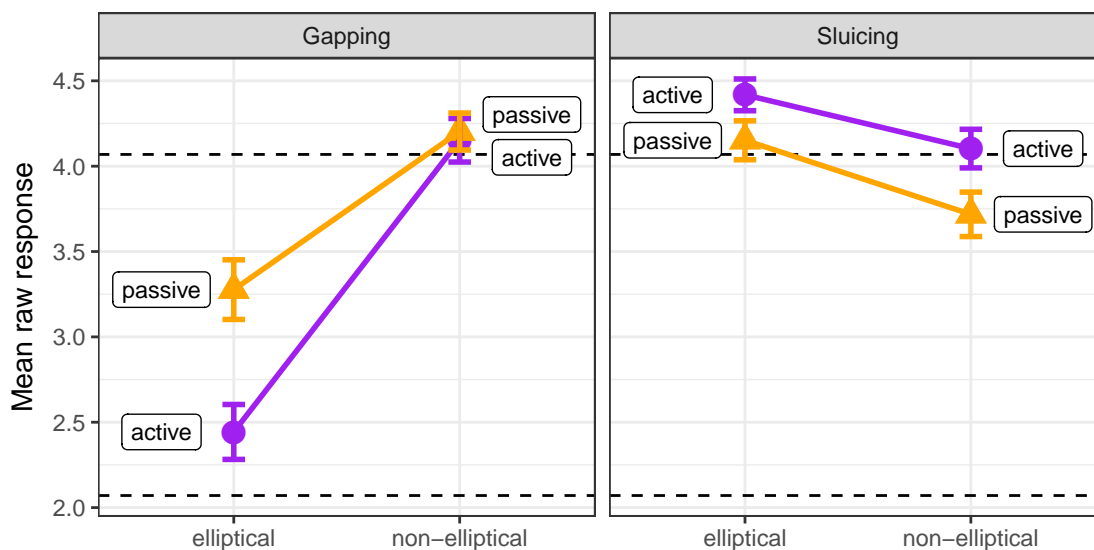


Figure 4: Results from pilot study for gapping (left) and sluicing (right). Errorbars reflect Standard Errors, dashed lines indicate mean ratings of (un)acceptable elliptical fillers.

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

Second, it is clear that the different forms of ellipsis display varied patterns of behavior, ones that may ultimately be tied to information structural factors that are specific to the linguistic properties of the respective forms. For instance, the robust penalty against no-ellipsis clauses found in the sluicing condition is likely due to a "repeated-clause" penalty, associated with the overt expression of a clause that could have been felicitously sluiced due to a representation of it being readily recoverable from the context. On the other hand, for gapping we found that ellipsis had a large, negative effect on acceptability. This negative effect might result from the fact that gapping sentences tend not to occur "out-of-the-blue," but instead only when an 'open proposition' (or equivalently, a multi-Wh QUD) exists in the context or can be accommodated from it (Sag 1976; Wilson & Sperber 1979; Prince 1986; Steedman 1990).¹³ As such, the effect may plausibly owe to the fact that the experimental stimuli failed to provide the sort of contexts necessary for gapping to be fully felicitous.

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

6 Conclusion

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

We have provided the results of three experiments that explored the source of the Mismatch Asymmetry, with particular attention to the predictions of the RH. Experiment 1 replicated the asymmetry, but also revealed that a parallel penalty occurs for syntactically matched cases in the passive voice as well. The results therefore suggest that the source of the penalty is more general

¹³ For example, Steedman (1990) says: "Indeed, even the most basic gapped sentence, like *Fred ate bread, and Harry, bananas*, is only really felicitous in contexts which support (or can accommodate) the presupposition that the topic under discussion is *Who ate what*." (p. 248).

than the domain over which the RH applies. Experiment 2 examined whether similar penalties are witnessed in variants in which there is no ellipsis. Consistent with the RH, the penalty did not generalize to the unelided variants. Experiment 3 provided a critical test of the theory by employing variants in which VP-ellipsis refers cataphorically. Whereas the RH predicts that cases with active voice ellipsis clauses should be rated as less acceptable than those with passive voice ellipsis clauses, the results revealed the opposite effect. In total, the results of Experiments 1-3 reveal consistent evidence for a penalty against ellipsis clauses in the passive voice. We therefore conclude that the explanation offered by the RH fails to explain the data, and hence that it provides no evidence, either for or against, syntactic theories of VP-ellipsis.

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

Supplementary files

All experimental stimuli, data, and analysis scripts for reproducing the results reported in this paper are available at <https://github.com/tpoppels/poppels-kehler-2019-glossa/>.

Funding information

We gratefully acknowledge support from NSF grant BCS-1456081.

Acknowledgments

The results of Experiments 1-3 were presented at the 31st Annual CUNY Sentence Processing Conference (CUNY 2018), the 54th Meeting of the Chicago Linguistics Society (CLS 2018), and the California Meeting on Psycholinguistics (CAMP 2018). We thank the audiences at those meetings as well as Brian Dillon, Helena Aparicio, Chuck Clifton, and Lyn Frazier for providing insightful feedback and discussion. We are also grateful to three anonymous reviewers and the special issue editors for helpful comments on an earlier draft of the paper.

References

- Arregui, Ana, Charles Clifton, Lyn Frazier & Keir Moulton. 2006. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language* 55(2). 232–246.
- Barr, Dale J, Roger Levy, Christoph Scheepers & Harry J Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278.

- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. <http://dx.doi.org/10.18637/jss.v067.i01>.
- Chao, Wynn. 1988. *On ellipsis*. New York: Garland.
- Dalrymple, Mary, Stuart M. Shieber & Fernando Pereira. 1991. Ellipsis and higher-order unification. *Linguistics and Philosophy* 14(4). 399–452.
- Frazier, Lyn. 2013. A recycling approach to processing ellipsis. In Lisa Lai-Shen Cheng & Norbert Corver (eds.), *Diagnosing syntax*, 485–501.
- Givón, Talmy. 1990. *Syntax: a functional typological introduction*. Amsterdam: John Benjamins.
- Grant, Margaret, Charles Clifton & Lyn Frazier. 2012. The role of non-actuality implicatures in processing elided constituents. *Journal of Memory and Language* 66(1). 326–343.
- Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein. 1995. Centering: a framework for modelling the local coherence of discourse. *Computational Linguistics* 21(2). 203–225.
- Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.
- Hardt, Daniel. 1990. *A Corpus-Based Survey of VP Ellipsis*. Unpublished Manuscript, University of Pennsylvania.
- Hardt, Daniel. 1999. Dynamic interpretation of verb phrase ellipsis. *Linguistics and Philosophy* 22(2). 187–221.
- Kehler, Andrew. 1993. The effect of establishing coherence in ellipsis and anaphora resolution. In *Proceedings of the 31st conference of the Association for Computational Linguistics (ACL-93)*, 62–69. Columbus, Ohio.
- Kehler, Andrew. 2000. Coherence and the resolution of ellipsis. *Linguistics and Philosophy* 23(6). 533–575.
- Kehler, Andrew. 2002. *Coherence, reference, and the theory of grammar*. CSLI Publications Stanford, CA.
- Kehler, Andrew. 2017. Opportunities and challenges for a referential theory of VP-ellipsis. Talk presented at the Workshop on Experimental and Corpus-based Approaches to Ellipsis. LSA Institute, Lexington, KY.
- Kehler, Andrew. 2019. Ellipsis and discourse. In Jeroen van Craenenbroeck & Tanja Temmerman (eds.), *Handbook of ellipsis*, 314–341. Oxford University Press.
- Kertz, Laura. 2013. Verb phrase ellipsis: The view from information structure. *Language* 89(3). 390–428.
- Kim, Christina S., Gregory M. Kobele, Jeffrey T. Runner & John T. Hale. 2011. The acceptability cline in VP ellipsis. *Syntax* 14(4). 318–354. <http://dx.doi.org/10.1111/j.1467-9612.2011.00160.x>.
- Kim, Christina S. & Jeffrey T. Runner. 2018. The division of labor in explanations of verb phrase ellipsis. *Linguistics and Philosophy* 41(1). 41–85.
- Lobeck, Anne. 1999. VP ellipsis and the minimalist program: Some speculations and proposals. In Shalom Lappin & Elabbas Benmamoun (eds.), *Fragments: studies in ellipsis and gapping*, 98–123. New York: Oxford University Press.
- Mehler, Jacques. 1963. Some effects of grammatical transformations on the recall of english sentences. *Journal of Verbal Learning and Verbal Behavior* 2(4). 346–351.
- Merchant, Jason. 2013. Voice and ellipsis. *Linguistic Inquiry* 44(1). 77–108.

- Miller, Philip & Geoffrey K. Pullum. 2014. Exophoric VP ellipsis. In Philip Hofmeister & Elizabeth Norcliffe (eds.), *The core and the periphery data-driven perspectives on syntax inspired by Ivan A. Sag*, 5–32. Stanford, CA.
- Parker, Daniel. 2017. Navigating ellipsis structures in memory: New insights from computational modeling. Talk presented at the Workshop on Experimental and Corpus-based Approaches to Ellipsis. LSA Institute, Lexington, KY.
- Prince, Ellen F. 1986. On the syntactic marking of presupposed open propositions. In *Papers from the parasession on pragmatics and grammatical theory at the 22nd regional meeting of the Chicago Linguistic Society*, 208–222. Chicago, IL.
- R Development Core Team. 2009. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <http://www.R-project.org>.
- Roberts, Craige. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5(6). Published version of draft circulated in 1996 and amended in 1998, 1–69. <http://dx.doi.org/10.3765/sp.5.6>.
- Rohde, Hannah & Andrew Kehler. 2014. Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience* 29(8). 912–927.
- Sag, Ivan A. 1976. *Deletion and logical form*. Massachusetts Institute of Technology dissertation.
- SanPietro, Steven A., Ming Xiang & Jason Merchant. 2012. Accounting for voice mismatch in ellipsis. In *Proceedings of the 30th West Coast Conference on Formal Linguistics*, 32–42. Somerville, MA: Cascadilla Proceedings Project.
- Schachter, Paul. 1977. Does she or doesn't she? *Linguistic Inquiry* 8(4). 762–767.
- Shibatani, Mashayoshi. 1985. Passives and related constructions: a prototype analysis. *Language* 61. 821–848.
- Steedman, Mark J. 1990. Gapping as constituent coordination. *Linguistics and Philosophy* 13(2). 207–263.
- Thoms, Gary. 2015. Syntactic identity, parallelism and accommodated antecedents. *Lingua* 166. 172–198.
- Van Craenenbroeck, Jeroen. 2012. Ellipsis, identity, and accommodation. Unpublished manuscript, KU Leuven. <http://jeroenvancraenenbroeck.net/s/paper-ellipsis-and-accommodation.pdf>.
- Webber, Bonnie Lynn. 1978. *A formal approach to discourse anaphora*. Reprinted in Outstanding Dissertations in Linguistics Series, Garland Publishers, 1979. Harvard University dissertation.
- Wilson, Deirdre & Dan Sperber. 1979. Ordered entailments: An alternative to presuppositional theories. In C. K. Oh & D. Dinneen (eds.), *Syntax and semantics XI: Presupposition*, 299–323. Academic Press.
- Xiang, Ming & Josef Klafka. 2018. Memory retrieval in comprehension is sensitive to production alternatives. 31st Annual CUNY Sentence Processing Conference. Davis, CA.