Acceptability, predictability and processing of antecedent-target mismatches under verb phrase ellipsis

Robin Lemke, Saarland University, DE, robin.lemke@uni-saarland.de

Deletion-based accounts of verb phrase ellipsis (VPE) predict that this construction requires a syntactically identical antecedent, but previous research shows that some antecedent-target mismatches are perceived as relatively acceptable in experiments (see e.g. Arregui et al., 2006; Miller & Hemforth, 2014). So far, the acceptability of these mismatches has been explained mostly by licensing conditions on VPE or by ellipsis-specific processing mechanisms. This article explores to what extent the acceptability of mismatches follows from the more general principles of an information-theoretic account of language use, which has been independently evidenced for other omission phenomena: To avoid under- or overutilizing the hearer's processing resources, predictable VPs are more likely to be omitted, whereas unpredictable ones are more likely to be realized. This hypothesis is tested with three experiments that investigate a gradual acceptability cline between VPE mismatches which has been reported by Arregui et al. (2006). First, an acceptability rating study replicates the overall pattern found by Arregui et al. (2006) and confirms that the effect is specific to ellipsis. Second, a production task shows that the acceptability differences are indeed related to a gradual decrease in the predictability of the target VP, which is also reflected in the likelihood of participants producing VPE. Finally, a self-paced reading experiment shows that VPE is more acceptable when it is easier to process. Overall, the experimental results support the information-theoretic account and suggest that no specific syntactic constraints or reconstruction mechanisms might be required to account for the acceptability cline observed for the mismatches investigated.

1. Introduction

Some deletion-based theories of ellipsis require elided expressions to have a syntactically identical antecedent, but previous research has identified several exceptions to this generalization. This mismatch between theoretical predictions and empirically observed data has been addressed by refining the licensing conditions on ellipsis so that some degree of (systematic) variation between antecedent and target is allowed: For instance, Chung (2006, 2013), Barros (2014) and Barros & Kotek (2019) define licensing conditions to account for antecedent-target mismatches under sluicing, and Johnson (2001) and Merchant (2013) do so for verb phrase ellipsis (VPE) (Sag, 1976; Williams, 1977). Refined licensing conditions may correctly capture the empirically observed acceptability patterns, but adding such constraints, which are often specific to a particular type of ellipsis, to the syntactic system increases its complexity. In contrast to this approach, some acceptability patterns in antecedent-target mismatches under ellipsis have been explained in terms of independently evidenced processing mechanisms. From this perspective, some mismatches are less acceptable, because they are more difficult to process. This article explores to what extent an information-theoretic processing account, similar to those successfully applied to other omission phenomena (see e.g. Jaeger, 2010; Kravtchenko, 2014; Kurumada & Jaeger, 2015; Lemke, 2021; Levy & Jaeger, 2007; Norcliffe & Jaeger, 2016), can explain an acceptability cline in antecedent-target mismatches under VPE.

In the literature on VPE, it is a well-established finding that VPE allows for some antecedent-target mismatches, whereas others appear to be heavily degraded. For instance, consider the case of voice mismatches (Kehler, 2002; Merchant, 2013): In (1a), where the material which is targeted by ellipsis (*give a book*) has a syntactically identical antecedent in the first conjunct, VPE is intuitively acceptable. In contrast, it is not in (1b), which contains a voice mismatch between a passive antecedent and an active target VP. However, (1c) shows that the complete picture is more complex: If voice mismatches were ungrammatical across the board, (1c) would be falsely predicted to be ungrammatical, too.

(1)   a.   John **gave Ann a book** and Sue did ⟨give Ann a book⟩, too.
        b.   *Ann was **given a book** by John and Sue did ⟨give Ann a book⟩, too.
        c.   This problem was to **have been looked into**, but obviously nobody did ⟨look into the problem⟩.

(Kehler, 2002, p. 548)

The acceptability of antecedent-target mismatches like (1b,c) is relevant to any theory of (VP) ellipsis for two reasons: First, the (un)acceptability of particular mismatches has been used as a diagnostic in the theoretical debate over the most appropriate analysis of VPE, and, second, an empirically appropriate theory of VPE must account for the observed acceptability pattern.

The syntactic theories of VPE proposed throughout the last 50 years can be roughly grouped into two camps: On the one hand, syntactic analyses assume that the ellipsis site contains unarticulated linguistic structure, which is deleted or silenced under some kind of identity relation with the antecedent (Kennedy, 2003; Merchant, 2013; Sag, 1976). On the other hand, pragmatic approaches analyze the ellipsis site as containing some kind of null anaphor, whose reference must be established pragmatically (Elbourne, 2008; Ginzburg & Sag, 2000; Hardt, 1993, 1999). As a rule of thumb, acceptable mismatches support the pragmatic view, according to which ellipsis is licensed as long as the omitted material is recoverable from context, no matter how (or whether) it is linguistically encoded. In contrast, unacceptable mismatches with salient antecedents support the syntactic perspective, according to which VPE requires a parallel linguistic antecedent. For instance, the contrast between (1a) and (1b) seems to require a syntactic explanation, because the mismatch is degraded, even though the antecedents are relatively meaning-equivalent, which makes the omitted VP easy to recover. The acceptability of (1c), in turn, supports a pragmatic account, because the mismatch is acceptable in the absence of an identical antecedent.

However, the data in (1) already show that neither of these perspectives is compatible with the data without additional assumptions. From the pragmatic perspective, it must be explained why mismatches like (1b) are degraded in spite of having a salient antecedent, and from the syntactic perspective, why mismatches like (1c) are acceptable, even though they lack a parallel antecedent. In previous research, theories of VPE have, therefore, been supplemented with (i) particular assumptions about syntactic structure (Johnson, 2001; Merchant, 2013), (ii) a reconstruction or accommodation mechanism for parallel antecedents (Arregui et al., 2006; Thoms, 2015; van Craenenbroeck, 2012), (iii) information- or discourse-structural licensing constraints on ellipsis (Grant et al., 2012; Hardt & Romero, 2014; Hendriks, 2004; Kehler, 2002; Kertz, 2013; Miller & Hemforth, 2014) or (iv) processing constraints (C. S. Kim et al., 2011). These mechanisms restrict the overgeneration of pragmatic accounts, which predict some unacceptable mismatches to be grammatical, or remedy the undergeneration of syntactic accounts, which predict acceptable mismatches to be ungrammatical.

Some of these approaches are specific to VPE (or even to particular types of VPE mismatches) and mainly motivated by the divergence between the empirically or introspectively observed data and theoretical assumptions. In this article, I explore whether an information-theoretic account of ellipsis usage, which is based on general processing principles, can explain the acceptability of mismatches under VPE. The central idea is that speakers tend toward reducing predictable expressions, which are easy to process, and encoding unpredictable ones more redundantly, in order to distribute the associated processing effort across more words. Applied to VPE, this predicts that ellipsis is more strongly preferred when the target VP is predictable in context. I hypothesize that it is the relatively high predictability of the target of ellipsis that makes

some mismatches acceptable. This account has two main advantages over VPE-specific accounts: First, it relies on an independently evidenced predictive processing strategy (Altmann & Kamide, 1999; Hale, 2001; Levy, 2008) and derives usage preferences from general mechanisms, which are not specific to ellipsis. Second, it models *optional* omissions and, therefore, predicts whether speakers actually choose to produce VPE (provided that it is licensed), whereas purely syntactic accounts only explain why VPE is available in a particular situation.

Section 2 of this article discusses previous approaches to VPE mismatches, and Section 3 presents the information-theoretic account. Its predictions are tested with three experiments: An acceptability rating study replicates the experiment by Arregui et al. (2006) and tests whether the effect is really ellipsis-specific (Section 4). A production task measures production preferences and tests whether the acceptability cline can be traced back to the predictability of the target VP (Section 5). Finally, a self-paced reading experiment investigates whether differences in predictability result in greater processing effort (Section 6). Section 7 investigates whether the experimental measures predict effects at the item level, which is expected under the information-theoretic account, but not under competing accounts, and Section 8 discusses the results of the experiments in the light of the different approaches to VPE mismatches.

2. Previous accounts of mismatch (un)acceptability

The (un)acceptability of presumably (un)grammatical antecedent-target mismatches under VPE has been approached with a wide range of syntactic and psycholinguistic accounts. In what follows, I briefly summarize the relevant proposals before I present an information-theoretic account of the phenomenon in Section 3. The approaches discussed in this section differ on various levels: (i) whether (some or possibly all) mismatches are grammatical, (ii) whether they make particular assumptions about the structure at the ellipsis site, (iii) whether the mechanism determining the (un)acceptability of mismatches is specific to (VP) ellipsis or more general, and (iv) whether it is pragmatic or processing-based. Some of the approaches are agnostic with respect to (ii), but others are designed to account for mismatches under certain theoretical assumptions.

2.1 Syntactic approaches

From a syntactic identity perspective, a possible explanation for acceptable mismatches is that they are only apparent mismatches on the surface, but that the underlying structure conforms to syntactic identity. This has been explicitly proposed for voice mismatches under VPE by Merchant (2013) and for VPE with nominal antecedents by Johnson (2001).

For the former case, Merchant (2013) argues that mismatches like (1b,c) are acceptable under VPE, because voice is encoded in a VoiceP dominating the target of VPE (which Merchant

identifies as *v*P), as (2) shows. Since the Voice head survives ellipsis, under this analysis, the elided constituent has an identical antecedent (3).[1]

(2) $[_{TP} \ldots [_{VoiceP} \ldots [_{vP} \ldots]]]$

(3) This problem was to have been $[_{VoiceP}$ Voice$_{passive}$ $[_{vP}$ Arg v look_into]], but obviously nobody$_2$ did $[_{VoiceP}$ Voice$_{active}$ $[_{vP}$ t_2 v look_into]].

(adapted from Merchant, 2013, p. 90)

Johnson (2001) proposes a similar analysis for acceptable instances of VPE with nominal antecedents, like (4). By contrasting the acceptable (4a) with the ungrammatical (4b), he argues that deverbal nouns are the result of syntactic derivation and that they contain a verbal head which can serve as an antecedent (unlike *infidelity* in (4b)). As Merchant (2013) assumes for voice mismatches, there is no actual syntactic mismatch involved in (4a), according to Johnson (2001).

(4) a. The candidate was dogged by charges of infidelity and [avoiding the draft], or at least trying to Δ.

(Johnson, 2001, p. 470, originally from Hardt, 1993, p. 35)

     b. *The candidate was dogged by charges of [infidelity], or at least trying to.

(Johnson, 2001, p. 470)

Merchant (2013) or Johnson (2001) predict categorical acceptability differences between grammatical apparent mismatches like (3) and (4a), on the one hand, and genuine mismatches, on the other hand: Only the latter should be ungrammatical, because they violate syntactic identity. More gradual acceptability patterns, like the one investigated in this article, therefore challenge these approaches, unless additional effects of processing, repair or discourse structure (which I discuss in what follows) are assumed. However, in that case, it would have to be shown that syntactic approaches have additional and independent explanatory power.

2.2 Parsing heuristics

C. S. Kim et al. (2011) explain observed acceptability contrasts among voice and category mismatches like (1b), (1c), and (4) with processing principles. Similarly to Merchant (2013) and Johnson (2001), C. S. Kim et al. (2011) also analyze these mismatches as, in principle, grammatical, but they argue that they can still be heavily degraded if they are hard to process, due to violations of parsing principles. C. S. Kim et al. (2011) investigate the effects of two parsing principles experimentally. First, they observe a tendency to omit the highest verbal

---

[1] In contrast, under sluicing, the complete TP, which includes VoiceP, is deleted. Since the antecedent and the target TPs differ in the Voice heads they contain, Merchant (2013) predicts voice mismatches under sluicing to be unacceptable:

(i) *Joe was murdered, but we don't know who. (Merchant, 2013, p. 81)

projection possible under identity (MaxElide, Merchant, 2008). Second, arguments seem to be aligned in canonical order, depending on their thematic role: complements of transitive verbs are expected to appear to the right of their heads. Their rating studies show that voice mismatches, which require higher verbal projections like VoiceP to survive ellipsis, are degraded as compared to matching instances of VPE, which allow for the omission of these projections and hence conform to MaxElide.

The account of C. S. Kim et al. (2011) makes less categorical predictions than syntactic accounts, since the penalties for violating heuristics can be additive. It should also be noted that they tested only for the potential effects of two principles, and the account could be extended to other acceptability contrasts if further heuristics are assumed.

2.3 Accommodation and reconstruction

In contrast to C. S. Kim et al. (2011), who analyze acceptable (apparent) mismatches as grammatical but possibly degraded, accommodation and reconstruction accounts view them as ungrammatical, but possibly acceptable when it is relatively easy to accommodate or reconstruct the missing antecedent.

Under an accommodation account (Thoms, 2015; van Craenenbroeck, 2012), VPE requires syntactic identity, but the identity constraint can be satisfied not only by overtly present antecedents, but also by accommodated antecedents. Which antecedents can(not) be accommodated is subject to further constraints: For instance, van Craenenbroeck (2012) argues that the accommodated antecedent cannot contain material which is focused or has been extracted from the ellipsis site and that it may include additional pronouns and copulas, which can be freely accommodated (Merchant, 2004). Under this view, mismatches can still conform to syntactic identity conditions, as long as accommodation is possible.

Arregui et al. (2006) take a similar approach, but they argue that the missing antecedent is syntactically reconstructed, instead of being accommodated. According to their VP recycling hypothesis, hearers apply syntactic derivation to the available linguistic material to construct a matching antecedent. The effort required for this varies as a function of the number of derivation steps required. Since great processing effort results in degraded acceptability (see e.g Hofmeister et al., 2013), VPE mismatches which involve more derivation steps are predicted to be more strongly degraded.

The main difference between their VP recycling hypothesis and the accommodation accounts is that the latter propose specific rules restricting what can be accommodated, whereas reconstruction is not guided by ellipsis-specific rules. Relying on more general syntactic mechanisms also distinguishes the VP recycling hypothesis from the processing account of C. S. Kim et al. (2011), who attribute the acceptability of mismatches to particular parsing heuristics like *MaxElide*.

Arregui et al. (2006) present data from a rating study which shows an acceptability decrease from (5a) through (5d), taken from Arregui et al. (2006, p. 234), in support of their account. The idea is that the number of derivation steps required to reconstruct a matching antecedent increases accordingly: In (5a), a VP identical to the omitted one is contained in the first conjunct, so there is no mismatch. In (5b), the VP is nominalized and embedded. In (5c), the subject *the comet* is additionally extracted out of the VP. Finally, in (5d), there is no verbal head of *to see* at all, but only an adjective. Consequently, VPE is predicted to be acceptable in (5a) and gradually degraded from (5b) (embedding) and (5c) (embedding + extraction) through (5d), where syntactic reconstruction is impossible.

(5)	a.	None of the astronomers saw the comet, but John did.	(Available VP)
	b.	Seeing the comet was nearly impossible, but John did.	(Embedded VP)
	c.	The comet was nearly impossible to see, but John did.	(VP with trace)
	d.	The comet was nearly unseeable, but John did.	(Negative adjective)

Even though Arregui et al. (2006) agree with C. S. Kim et al. (2011) on the relevance of parsing principles, the former assume that all of the mismatches in (5) are ungrammatical and that the repair mechanism is specific to VPE, whereas the latter view at least some mismatches as grammatical and resort to more general parsing heuristics.

2.4 Memory retrieval

Parker (2018) pushes in a similar direction as the VP recycling hypothesis by providing a processing-based explanation for the acceptability cline observed by Arregui et al. (2006). His proposal agrees with the VP recycling hypothesis in viewing VPE mismatches as ungrammatical, yet sometimes acceptable, but Parker (2018) derives the pattern from memory retrieval mechanisms, which are required to resolve ellipses and other filler-gap dependencies. As with other dependencies, he argues that the processing of ellipsis involves searching for an antecedent from working memory when the hearer encounters the ellipsis site (in the case of VPE, the auxiliary or modal). Antecedents which agree with the ellipsis site to a larger extent in terms of morphosyntactic properties, like voice or syntactic category, are more strongly activated in this search process. The higher activation facilitates their retrieval, which results in reduced processing effort and leads to higher acceptability. Parker (2018) supports his account of the acceptability cline with a computational model of working memory.

The central empirical prediction of the memory retrieval account with respect to VPE mismatches is that mismatches are more acceptable the more syntactically similar the antecedent and target are, because the activation of the antecedent in memory is stronger when it matches more features with the ellipsis site. This predicts the same pattern that Arregui et al. (2006) found, but the explanations differ. Furthermore, the quantitative predictions of both approaches can differ, depending on the amount and type of derivation steps and the set of distinctive features predicting activation in working memory which are assumed.

2.5 Information- and discourse-structural approaches

The accounts discussed so far predict and explain effects of the form of the antecedent and its similarity to the potential target of ellipsis, which will be also investigated in this article. However, there is experimental evidence that the acceptability of VPE mismatches is also determined by discourse, information structure and even extralinguistic context. These factors do not explain the acceptability cline reported by Arregui et al. (2006) (because there is no context manipulation in their experiments), but since context can affect the acceptability of mismatches, an empirically appropriate account of VPE mismatches should also take this into account.

Kehler (2000, 2002) emphasizes the importance of discourse relations between the conjuncts containing the antecedent and target for the acceptability of VPE mismatches: Mismatches are degraded when a connective like *and (. . . too)* suggests that a parallel relation holds between antecedent and target (6a), but not in the case of other discourse relations, like (6b), involving a cause-effect relation (more specifically, a violated expectation in Kehler's terminology).

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

(Kehler, 2000, p. 551)

     b. This problem was to **have been looked into**, but obviously nobody did ⟨look into the problem⟩.

(Kehler, 2000, p. 548)

Kertz (2013) reanalyzes the data in Kehler (2000, 2002) in terms of information structure: She argues that degraded VPE mismatches in Kehler's parallel relations contain contrastive topics, which raise the expectation of further parallel structure. Furthermore, Kertz finds similar effects for nonelliptical utterances, and concludes that mismatches are not degraded due an ellipsis-specific mechanism.

Miller and Hemforth (2014) find an effect of implicit Questions under Discussion (QuDs), which can be triggered by the antecedent, on the acceptability of VPE with a nominal antecedent. The contrast in (7) shows that VPE is more acceptable when the antecedent raises an implicit polar question which is picked up in the target of ellipsis (7a), than when it does not (7b).

(7)   a. The integrity of the Senate depends on her **participation**. If she does, […] (compare: depends on whether or not she participates)

(Miller & Hemforth, 2014, p. 7)

     b. That depends on her **answer**. (≈ That depends on what her answer is ≠ That depends on whether or not she answers) #If she does […]

(Miller & Hemforth, 2014, p. 8)

Similarly, Grant et al. (2012) show that the inclusion of non-actuality implicature triggers, like certain modals (8a), which raise the inference that the event described in the antecedent did not

occur, improves the acceptability of mismatches as compared to the same mismatches without the trigger (8b).[2]

(8)   a.   This information **needed to** be released but Gorbachev didn't.

                                                     (Grant et al., 2012, p. 331)

        b. ?This information was released but Gorbachev didn't.   (Grant et al., 2021, p. 329)

Finally, Geiger and Xiang (2021) show that the resolution of VPE is affected not only by linguistic context, but also by extralinguistic context, which they model with graphical stimuli consisting of comic strips. Their experiments do not explicitly deal with mismatches, but in some of their conditions, they tested instances of discourse-initial VPE, which clearly lack an overt linguistic antecedent. The data indicate that extralinguistic context is taken into account also when linguistic context is present, and not only as a last resort strategy applied in its absence.

2.6 Summary

The accounts of VPE mismatches discussed in this section differ along various dimensions, even though not all of the theories are committed to particular assumptions about each of these: (i) the grammaticality of mismatches, (ii) the content of the ellipsis site, (iii) ellipsis-specific or general mechanisms determining acceptability, and (iv) pragmatic vs. psycholinguistic mechanisms.

While the syntactic accounts discussed in 2.1 imply a grammaticality split between acceptable (grammatical) and ungrammatical mismatches, the other proposals do not make this distinction. The VP recycling hypothesis (Arregui et al., 2006) and the memory retrieval account (Parker, 2018) view all mismatches as ungrammatical, but possibly acceptable, whereas these mismatches are grammatical, but possibly degraded, according to the processing account of C. S. Kim et al. (2011). The information- and discourse-structural accounts are not committed to either of these perspectives.

The accounts also disagree with respect to the question of whether or not there is unarticulated structure at the ellipsis site. On the one hand, the syntactic, reconstruction and accommodation accounts, as well as the processing account of C. S. Kim et al. (2011), assume that there is unarticulated structure at the ellipsis site, which is what requires a syntactically identical (possibly reconstructed) antecedent. On the other hand, the information- and discourse-structural accounts are also compatible with the assumption that there is no structure or just a null anaphor at the ellipsis site. However, they do not speak against unarticulated syntactic structure, since an appropriate antecedent might also become salient through discourse or context.[3]

---

[2] The grammaticality judgment is mine, since Grant et al. (2012) did not include any in their paper, but find that the (8b) is degraded.

[3] This has been proposed by Merchant (2004) for fragments.

Beyond the question of what the ellipsis site looks like and whether acceptable mismatches are grammatical, the accounts propose different mechanisms to account for their acceptability. The information- and discourse-structural approaches operate on pragmatic concepts, whereas the accounts based on memory retrieval, processing constraints and reconstruction rely on processing mechanisms. Additionally, the reconstruction accounts propose a mechanism specific to VPE, whereas the others rely on more general principles.

From an empirical perspective, the studies discussed in this section show that the acceptability of mismatches is affected not only by syntactic factors, but also by linguistic and extralinguistic context. This suggests that listeners regularly take into account different sources of information when processing VPE. Purely syntactic theories of VPE mismatches or those assuming a syntactic reconstruction mechanism cannot account for this observation.

3. An information-theoretic account

The discussion of previous studies on VPE mismatches shows that an account of this phenomenon needs to consider structural properties of antecedent and target, like morphosyntactic similarity (Arregui et al., 2006; C. S. Kim et al., 2011; Parker, 2018), as well as extralinguistic and pragmatic factors, like discourse- and information-structure-based expectations (Grant et al., 2012; Kertz, 2013; Miller & Hemforth, 2014). Furthermore, the experimental studies by e.g. Arregui et al. (2006) and C. S. Kim et al. (2011) suggest that the data to be explained are inherently gradual, so a binary distinction between grammatical and ungrammatical mismatches seems to be insufficient. The gradual differences in VPE acceptability have been related to processing effort (Arregui et al., 2006; C. S. Kim et al., 2011; Parker, 2018).

The relationship between the predictability of the potential target of VPE, the processing effort associated with ellipsis resolution and a gradual preference for ellipsis is very much expected under an information-theoretic approach to the usage of reduction phenomena. According to such accounts, expressions which are predictable from context are more likely to be reduced, whereas more explicit encodings are chosen for unpredictable expressions. Since predictability has been shown to index processing effort (Demberg & Keller, 2008; Hale, 2001; Levy, 2008), this strategy avoids under- or overutilizing the hearer's processing resources (Fenk & Fenk, 1980) and, thus, results in more efficient communication. Empirical research supports the expected relationship between predictability and reduction at different levels of linguistic analysis, such as phonological (Aylett & Turk, 2004) and morphological (Frank & Jaeger, 2008) reduction, the omission of function words and grammatical markers (Jaeger, 2010; Kurumada & Jaeger, 2015; Levy & Jaeger, 2007; Norcliffe & Jaeger, 2016) as well as content words (Kravtchenko, 2014; Lemke, 2021), pronominalization (Tily & Piantadosi, 2009), the shortening of word forms (Mahowald et al., 2018) as well as VPE with matching antecedents (Schäfer et al., 2021) and antecedent-target mismatches under sluicing (Lemke et al., 2022).

The account proposed in this article relies on two findings of the previous research on information-theoretic constraints on language use: First, speakers optimize their utterance so that processing it neither under- nor overutilizes the hearer's processing resources, and, second, speakers use ellipsis to achieve this goal. The idea of uniformly distributing processing effort can be traced back to the Constant Flow of Information principle (Fenk & Fenk, 1980) and has been more recently picked up in the related Smooth Signal Redundancy (Aylett & Turk, 2004) and Uniform Information Density hypotheses (Levy & Jaeger, 2007). Since predictable words are easier to process (Demberg & Keller, 2008; Hale, 2001; Levy, 2008), the processing effort caused by a word is indexed by its Shannon information, or Surprisal, which is calculated as the negative logarithm of its likelihood to occur in context (Levy, 2008; Shannon, 1948):

$$Surprisal(word \mid context) = -\log_2 p(word \mid context) \tag{1}$$

In order to communicate efficiently, speakers try to transmit as much information per unit of time as possible, without exceeding the (assumed) maximum capacity of the hearer's processing resources. Among other strategies (e.g. reordering words; Cuskley et al., 2021), they can make use of the optional reduction or insertion of linguistic material: On the one hand, speakers can reduce or omit predictable expressions, which would make an inefficient use of the hearer's processing resources otherwise. On the other hand, encoding unpredictable expressions more redundantly can avoid temporarily exceeding the hearer's capacity. Applied to VPE, these strategies predict a stronger tendency to reduce, i.e. omit, predictable VPs.[4] As the left panel of **Figure 1** suggests,
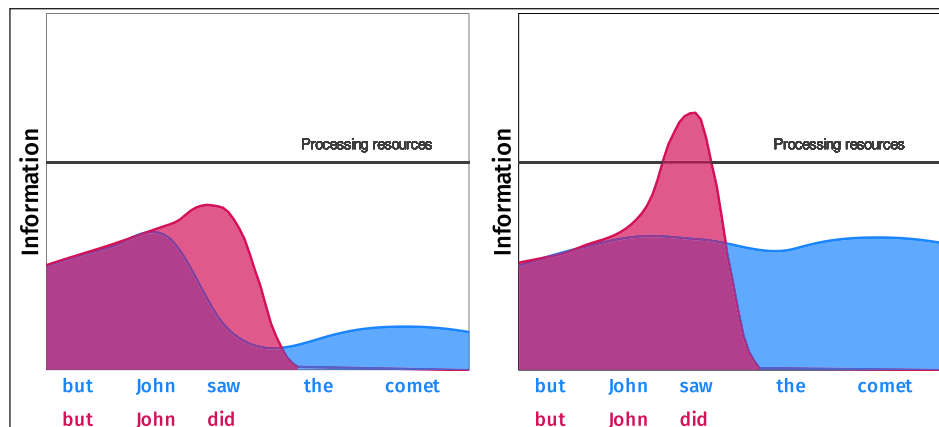

Figure 1: Hypothetical information density profiles for instances of predictable (left) and unpredictable (right) VPs and the corresponding VPE. The information density profile for the elliptical utterances is illustrated by the pink areas, whereas the blue ones represent the distribution of information across the nonelliptical utterances.

---

[4] VPE is not the only option, since VPs can also be reduced to *do so/do it* anaphora (Hankamer & Sag, 1976). Since there are partially different usage restrictions on these constructions than on VPE, I do not consider this option further in this article.

this avoids inefficient *troughs* in the information density profile of the utterance. In contrast, reducing an unpredictable expression can lead to a *peak* in the information density profile, which exceeds the hearer's processing resources. Inserting additional redundancy, i.e. using a complete VP, distributes information across more words and time, so that the peak is smoothed (see the right panel of **Figure 1**).

The notion of *predictability* used in what follows is not equivalent to the likelihood of actually encountering a word in a corpus in a particular context: If predictable expressions are more often reduced, extremely predictable expressions might even be very rarely explicitly realized, just because of their predictability.[5] Predictability, therefore, can be thought of as the likelihood of a particular expression appearing at some point in the sentence at a semantic level in the context of the preceding material – independently of its overt realization. In the case of VPE, this context consists of the first conjunct, the subject of the second conjunct, as well as preceding utterances (if there are any) in the discourse and extralinguistic context. For instance, world knowledge makes the omitted VP in (9a) probably more predictable than the VP in (9b), because cats love to sleep in beds, but people rarely catch mice.

(9)   a.   Peter slept in his bed and the cat did, too.
        b.   The cat caught a mouse and Peter did, too.

Equation 2 shows how the predictability of the target VP in coordination structures like (9) is calculated, given three possible continuations of the second conjunct after the subject *the cat/ Peter*: (i) the complete VP suggested by the first conjunct, (ii) another complete VP (*climbed on a tree*, *ordered a pizza*, etc.), and (iii) VPE.[6] High ratios of VPE and identical continuations increase predictability, whereas high ratios of other VPs reduce it.

$$p(VP_i|context) = \frac{p(explicit\ VP_i|context) + p(VPE|context)}{p(explicit\ VP_i|context) + p(VPE|context) + p(explicit\ VP_{\neg i}|context)} \quad (2)$$

With respect to antecedent-target mismatches under VPE, the information-theoretic account makes the same prediction as for ellipsis in general: Reducing an expression – in this case, a VP, which is maximally reduced by ellipsis – is more strongly preferred, the more redundant, i.e. predictable, this VP is. Consequently, a mismatch is expected to be relatively acceptable when the target is predictable, and degraded when it is not. Since predictability depends on linguistic factors like information structure (Kertz, 2013) and extralinguistic factors (Geiger & Xiang, 2021), both are expected to affect the acceptability of ellipsis, including antecedent-target mismatches.

---

[5] If listeners are aware of this, the overt realization might even be unexpected, because the likelihood of VPE increases with the probability of the target. I return to this issue in Section 4.

[6] This is simplified for expository purposes, since other omissions and reductions (e.g. *do so* anaphora or (pseudo) gapping) could also be used.

The information-theoretic account could also explain the acceptability cline reported by Arregui et al. (2006), if there were a similar decrease in VP predictability across the four conditions in (5). Unlike the VP recycling hypothesis, which relies on a mechanism specific to a particular type of ellipsis, the information-theoretic account would explain the data with general processing principles. This would provide a unified account of VPE mismatches and other omission and reduction phenomena, including ellipsis without antecedent-target mismatches, which have been shown to be subject to information-theoretic constraints.

If a gradual decrease in VP predictability across conditions were empirically confirmed, the empirical predictions of the information-theoretic account with respect to the overall pattern in the acceptability cline reported by (Arregui et al., 2006) would be aligned with their VP recycling hypothesis (Arregui et al., 2006) and the memory retrieval account (Parker, 2018). However, its predictions are more specific with respect to the production of mismatches and to item level effects.

The VP recycling hypothesis and the memory retrieval account view mismatches as ungrammatical, whereas the information-theoretic account implies that those mismatches which are produced by speakers on purpose (leaving aside errors) are grammatical: Information-theoretic frameworks assume, by definition, a parallel parser (Hale, 2001; Levy, 2008), which is restricted to the set of utterances which can be derived by applying the grammatical rules of a language. Utterances that do not conform to these rules are not considered by speakers as possible encodings of their message and should not be produced on purpose. Therefore, if subjects produced VPE mismatches and the ratio of VPE mismatches decreased in line with the target VP's predictability, this would particularly support the information-theoretic account.

The information-theoretic account also predicts more fine-grained differences between items, because the predictability of the target VP, which is expected to determine acceptability, might differ among individual stimuli. In contrast, the VP recycling hypothesis and the memory retrieval account derive acceptability from the degree of morphosyntactic similarity between antecedent and target directly. Since this similarity is kept constant across the stimuli within a condition, no systematic differences between stimuli are expected. Therefore, such effects will provide a further testing ground to distinguish between the proposals.

In what follows, I present three experiments which test the predictions of the information-theoretic account and which rely on the stimuli provided by Arregui et al. (2006) in the appendix to their article. First, an acceptability rating task aims at replicating the pattern they found for elliptical utterances, but also tests the nonelliptical counterparts. This ensures that possible differences in acceptability are ellipsis-specific and do not result from a general penalty for mismatching conjuncts (Kertz, 2013; C. S. Kim & Runner, 2011). Second, I use a production task to measure the predictability of the potentially reduced VP and actual production preferences.

This is crucial for evaluating the information-theoretic account, because it assumes that differences in predictability of the VP are the main reason for its reduction. Finally, a combined self-paced reading and speeded acceptability judgment experiment investigates how difficult the processing of ellipsis is across conditions, and whether processing effort is related to acceptability. Since information-theoretic processing accounts trace processing effort back to predictability differences, I expect that VPE is more difficult to process when the VP is less predictable and VPE less acceptable. I first present the three experiments independently (Sections 4–6) and investigate whether the data are related at the item level in the way predicted by the information-theoretic account in Section 7.

4. Experiment 1: Acceptability rating study

Experiment 1 uses a web-based acceptability rating task to replicate the data of Arregui et al. (2006) and to investigate whether the acceptability cline they report is really specific to ellipsis. Arregui et al. (2006) conducted two experiments using stimuli like (10) ($n = 16$), one using a binary acceptability judgment task and the other using a 5-point Likert rating scale. In both experiments, they find evidence for a gradual acceptability cline from (10a), with a syntactically identical antecedent VP, through (10d), with no verbal head at all in the first conjunct. In order to distinguish between the effects of VPE mismatches and ellipsis-independent properties of the materials, they additionally tested nonelliptical controls consisting of the first conjunct of (10a–d) only (e.g. *None of the astronomers saw the comet*). Since the controls do not exhibit the acceptability cline evidenced for ellipsis, Arregui et al. (2006) conclude that the effect is indeed driven by ellipsis.

(10)   a.   None of the astronomers saw the comet, but John did.   (Available VP)
       b.   Seeing the comet was nearly impossible, but John did.   (Embedded VP)
       c.   The comet was nearly impossible to see, but John did.   (VP with trace)
       d.   The comet was nearly unseeable, but John did.   (Negative adjective)

In the replication of their experiment, I used the VPE conditions of their materials (e.g. (10)), which they provide in the appendix to their article. In addition, I tested the nonelliptical counterparts of the materials in (10), i.e. by inserting the omitted VP in the second conjunct (as in (11), for (10a)). This allows for one to assess whether the observed pattern is really ellipsis-specific, which is necessary if one is going to base any theory of mismatches under ellipsis on it. Previous empirical research found that structural mismatches between conjuncts are also degraded without ellipsis (Kertz, 2013; C. S. Kim & Runner, 2011), therefore, the possibility that the pattern found by Arregui et al. (2006) simply results from a general penalty for differing conjuncts needs to be ruled out.

(11)   None of the astronomers saw the comet, but John saw the comet.

Unlike other theories of antecedent-target mismatches under VPE, the information-theoretic account makes predictions about the acceptability of the nonelliptical controls depending on the likelihood of the target VP, because it is not specific to ellipsis: If the VP were extremely predictable, its overt realization would be highly redundant and, consequently, strongly dispreferred. If hearers assume that the speaker behaves rationally, they will expect ellipsis under these circumstances,[7] so that an overt identical VP might even become harder to process, because it is unexpected. However, as Experiment 2 will show, in the case of the stimuli tested here, the target VP does not seem to be extremely likely in any of the conditions.

4.1 Materials

The materials were identical to those used by Arregui et al. (2006) in the ellipsis conditions (10). In the nonelliptical conditions, the auxiliary *did* was replaced by an overt VP (11).

4.2 Procedure

The web-based acceptability rating study was conducted using the LimeSurvey presentation software (LimeSurvey GmbH, 2012). 64 self-reported native speakers of American English were recruited on the crowdsourcing platform Prolific (http://prolific.co) and rewarded with £2.50 for participating. Participants were asked to rate the naturalness of each of the sentences on a 7-point Likert scale with labeled endpoints (1 = completely unnatural, 7 = completely natural). They were assigned to one of eight lists, in which the materials were distributed so that each participant saw each of the 16 token sets once and each of the four ANTECEDENT conditions in (10) equally often ($n = 4$). COMPLETENESS (ellipsis/complete) was tested as a between-subjects variable in order to keep the ellipsis study comparable to the one by Arregui et al. (2006). The stimuli were mixed with 60 fillers, which included instances of gapping, sentences with possible subject gaps and garden path sentences with reduced complement and relative clauses. These materials were presented in individual pseudo-randomized order, ensuring that no two stimuli of the same type (e.g. items of Experiment 1 or gapping fillers) immediately followed each other.

4.3 Results

Figure 2 provides an overview of the mean ratings across the eight conditions. The data were analyzed with Cumulative Link Mixed Models for ordinal data (Christensen, 2022) in R, using a backward model selection procedure: Starting from a full model containing fixed effects for all predictors and their interactions as well as the maximal random effects structure supported by the data, I subsequently removed fixed effects that did not significantly improve model fit (starting with the interactions), as evidenced by Likelihood Ratio tests calculated using the anova

---

[7] See Yoshida et al. (2013) and N. Kim et al. (2020) for empirical evidence that under some circumstances, comprehenders expect a sentence they are processing to be elliptical (provided that the ellipsis is syntactically licensed).

function (R Core Team, 2022), to find the simplest model capable of appropriately explaining the data. The *p*-values reported for the final model were also obtained with this method, that is, by comparing a model containing the predictor in question to a model without it. This model selection approach was used in all analyses reported in this article.
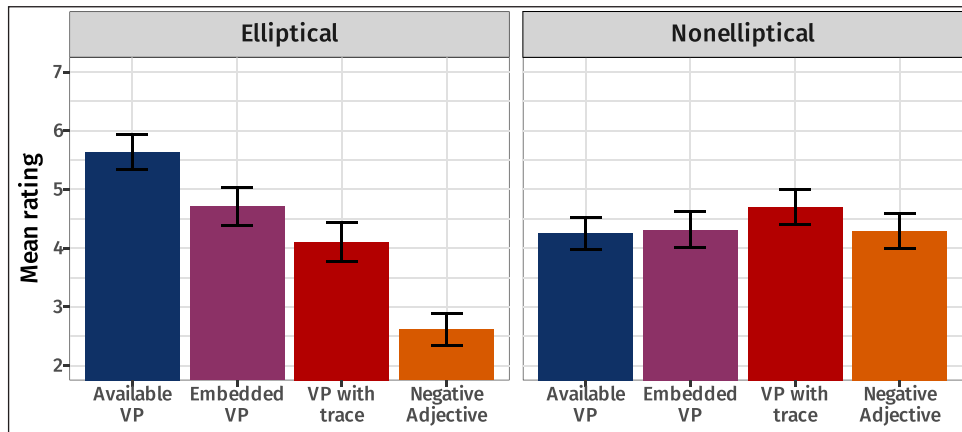

Figure 2: Mean ratings for the eight conditions tested in Experiment 1. Error bars indicate 95% confidence intervals.

In order to keep the analysis comparable to Arregui et al. (2006), I first analyzed the data sets for VPE and nonelliptical controls separately. I then conducted an analysis of the complete data set, in order to test for interactions between ANTECEDENT and COMPLETENESS, which would evidence effects that are ellipsis-specific. COMPLETENESS was deviation-coded (ellipsis = 0.5, complete VP = –0.5), whereas I used forward coding for ANTECEDENT, in order to test the hypothesis of a gradual acceptability decrease from (11a) to (11d): As **Table 1** shows, the factor was transformed to three variables comparing (i) Available VP vs. the rest, (ii) Available/embedded VP vs. VP trace/Neg. Adjective, and (iii) Neg. Adjective vs. the rest. In what follows, I use the labels in **Table 1** to refer to these predictors. In addition to the predictors of theoretical interest, the scaled and centered POSITION of the trial in the experiments was considered in the analyses to quantify and factor out potential familiarization effects with the materials or the task.

Table 1: Forward coding scheme for ANTECEDENT.

Predictor	Available VP	Embedded VP	VP with trace	Neg. adjective
AVP v EVP	3/4	–1/4	–1/4	–1/4
EVP v TRACE	1/2	1/2	–1/2	–1/2
TRACE v ADJ	1/4	1/4	1/4	–3/4

Table 2: Fixed effects in the final model for the elliptical rating data.

Predictor	Estimate	SE	χ^2	p
AVP v EVP	1.71	0.42	13.42	<0.001
EVP v TRACE	0.92	0.46	3.75	$=0.053$
TRACE v ADJ	2.17	0.32	23.99	<0.001
POSITION	0.11	0.13	0.68	>0.4
AVP v EVP:POSITION	–0.6	0.24	6.15	<0.05

4.3.1 VPE data set analysis

The full model for the VPE analysis contained fixed effects for ANTECEDENT (forward-coded, see **Table 1**), the scaled and centered POSITION of the trial in the experiment, and all two-way interactions, as well as by-item and by-subject random intercepts and random slopes for ANTECEDENT and POSITION.[8] The final model is summarized in **Table 2**. Two out of three contrasts between the experimental conditions are significant: The available VP condition is significantly more acceptable ($\chi^2 = 13.42$, $p < 0.001$) and the negative adjective condition is significantly less acceptable than the other ones ($\chi^2 = 23.99$, $p < 0.001$). The medial contrast between the embedded VP condition and the VP with Trace condition is marginal ($\chi^2 = 3.75$, $p = 0.053$). In sum, these results replicate the pattern found by Arregui et al. (2006). The POSITION of the trial in the experiment has no significant effect on acceptability itself ($\chi^2 = 0.68$, $p > 0.4$), but there is a significant interaction between POSITION and the AVP v EVP contrast ($\chi^2 = 6.15$, $p < 0.05$), which suggests familiarization with the degraded mismatch conditions in the course of the experiment.

4.3.2 Complete data set analysis

In order to quantify the differences between elliptical and nonelliptical conditions, an analysis of the complete data set (including the nonelliptical data) additionally included COMPLETENESS as a binary predictor. The full model contained fixed main effects for the ANTECEDENT predictors, COMPLETENESS, the scaled and centered POSITION of the trial in the experiment and all two-way interactions between these predictors. It had by-subject and by-item random intercepts and random slopes for the ANTECEDENT predictors, POSITION, and additional by-item random

---

[8] Rating ~ (AVP v EVP + EVP v Trace + Trace v Adj) * Position + (1 + AVP v EVP + EVP v Trace + Trace v Adj + Position | Subject) + (1 + AVP v EVP + EVP v Trace + Trace v Adj + Position | Item).

slopes for COMPLETENESS and its interactions with the three ANTECEDENT predictors (since COMPLETENESS was tested between subjects, including related by-subject effects was not reasonable).[9]

Table 3: Fixed effects in the final model for the complete rating data.

Predictor	Estimate	SE	χ^2	p
COMPLETENESS	–0.21	0.47	0.2	> 0.6
AVP v EVP	0.84	0.25	9.41	< 0.01
EVP v TRACE	0.18	0.23	0.64	> 0.4
TRACE v ADJ	1.58	0.22	24.93	< 0.001
AVP v EVP:COMPLETENESS	1.77	0.42	16.31	< 0.001
EVP v TRACE:COMPLETENESS	1.79	0.38	19.47	< 0.001
TRACE v ADJ:COMPLETENESS	1.67	0.37	19.58	< 0.01

The final model is summarized in **Table 3**. The significant interactions of all three ANTECEDENT predictors with COMPLETENESS show that the patterns for VPE and nonelliptical utterances differ significantly from each other. Beyond the interactions, the main effects of the ANTECEDENT contrasts exhibit a similar pattern to the ellipsis data, except for the nonsignificant EVP v TRACE contrast, which is probably due to the high acceptability of the VP trace condition in the nonelliptical data. Overall, VPE is not particularly preferred ($\chi^2 = 0.2$, $p > 0.6$) and there was no significant main effect of, or interaction, with POSITION.

4.4 Discussion

Experiment 1 replicated and extended the second experiment of Arregui et al. (2006), using a different statistical analysis technique; it confirms that the gradual acceptability cline they report is ellipsis-specific. In addition to Arregui et al. (2006), who showed that the acceptability of VPE is not driven by the acceptability of the first conjunct alone, my data suggest that it cannot be explained by the interaction between the first and the complete second conjunct. The data for the nonelliptical controls do not show the gradual pattern found for the ellipsis conditions, which is reflected in highly significant CONDITION:COMPLETENESS interactions for all three

---

[9] Rating ~ (AVP v EVP + EVP v Trace + Trace v Adj) * Position + (AVP v EVP + EVP v Trace + Trace v Adj) * Completeness + Completeness:Position + (1 + AVP v EVP + EVP v Trace + Trace v Adj + Position | Subject) + (1 + AVP v EVP + EVP v Trace + Trace v Adj + Completeness | Item).

contrasts. The next two experiments turn to investigating whether the effect observed for the ellipsis conditions can be explained by the information-theoretic account.

5. Experiment 2: Production study

Experiment 2 used a production task to investigate two main predictions of the information-theoretic account. First, given the acceptability cline observed in the rating data, it is expected that the predictability of the potentially omitted VP increases in line with the acceptability of VPE. Since cloze probability is correlated with processing effort (Smith & Levy, 2011), this would reflect a strategy to distribute processing effort uniformly by (i) reducing predictable material in order to avoid underutilizing the hearer's processing resources, and (ii) increasing the redundancy in the encoding of unpredictable material, so that exceeding the processing resources is also avoided. Second, the information-theoretic account predicts that differences in the predictability of the potentially reduced VP will affect production preferences: If the VP is more predictable in a condition, speakers should relatively more often *produce* an instance of VPE in the production task instead of an explicit VP.

5.1 Materials

The materials were based on the stimuli of Arregui et al. (2006), tested in Experiment 1. For the production task, the materials were cut off after the subject of the second conjunct, as shown in (12) for the sample item in (5).

(12) a. None of the astronomers saw the comet, but John _____ (Available VP)
b. Seeing the comet was nearly impossible, but John _____ (Embedded VP)
c. The comet was nearly impossible to see, but John _____ (VP with trace)
d. The comet was nearly unseeable, but John _____ (Negative adjective)

5.2 Procedure

The production study was conducted over the Internet, using the LimeSurvey presentation software. 120 self-reported native speakers of American English, recruited on Prolific, participated in the experiment. Subjects who had completed the rating study were blocked from participating. Each participant was rewarded with £2.50. Subjects were asked to provide the most natural continuation to stimuli like (12). They were explicitly instructed not to be creative or funny, but to produce the continuation that they considered to be the most likely one. This was intended to counterbalance the possibility that participants might actually avoid highly predictable continuations, because they did not consider them interesting or newsworthy. The materials were distributed across eight lists, so that each participant saw each of the 16 token sets once and each of the four conditions equally often.[10] The 16 stimuli were combined with 16

---

[10] Two of each of these lists differed only in the materials of the other experiment that they were presented together with.

items from another experiment and presented in individual fully randomized order. The items from the other study also consisted of conjoined sentences to be completed after the subject of the second conjunct.

5.3 Annotation

The production data were annotated for a series of variables on which the predictors in the statistical analyses reported below were based. These variables concern three properties of the participants' responses: first, whether the continuation produced is identical to the VP expected, given the antecedent, i.e. *saw the comet* in the case of (12); second, whether this VP was complete or somehow reduced (and, if so, how); and, third, whether the VP appeared immediately after the subject of the second conjunct. Responses which consisted in obvious errors or which were not meaningful continuations of the utterance were excluded before annotation. This involved 4 data points (0.2% of the complete data).

5.3.1 Parallelism

The first annotation layer tracked whether the VP expected given the antecedent, to which I refer as *parallel* in what follows, was included in a continuation of the sentence (13a) or whether the participant completed the sentence in a different fashion, as in e.g. (13b). Reduced forms of a parallel VP, i.e. VPE, *do it, do so* and *do that* anaphora were also classified as parallel, because they require an antecedent, which is available only from the first conjunct. Continuations which are similar to a parallel one, but do not imply the parallel VP, were categorized as not parallel: (13c) neither implies that John saw the comet nor that he pictured it. When determining whether to classify a VP as parallel, I disregarded possible postverbal VP adjuncts, like *with his telescope* in (13d), and I also did not distinguish between VPs immediately following the subject and those appearing after further material, such as the modal *could* in (13d) (on this point, see 5.3.3).

(13)   a.   . . . saw the comet
         b.   . . . brought his telescope
         c.   . . . took some pics through his old telescope
         d.   . . . could see it with his telescope

For all VPs classified as parallel, it was annotated whether the lexical material they contain is strictly identical to the expected VP (*saw the comet*) or whether it contains synonyms, such as e.g. *spot* instead of *see*.[11] This annotation layer allows one to isolate definitive possible sources for

---

[11] It should be noted that the term *synonym* in this sense does not imply that the words were genuine synonyms independently of context, but that they refer to the same entity or action in the context of the stimulus and continuation produced by participants. For instance, even though *to beat* and *to defeat* are not synonyms in general, in the context

VPE (fully identical target VPs), but also to take into account the possibility of subjects avoiding the repetition of lexical material for stylistic reasons, even though they intend their VP to be meaning-equivalent to the antecedent VP. Since it is not always certain whether participants resorted to the synonym for stylistic reasons or whether they intended a change in meaning, I will rely on the stricter notion of identity (treating synonyms as non-parallel) in the main analysis below, but also show that – at least in my data set – the results are similar for the weaker notion of identity (including synonyms).

5.3.2 Reduction

For all parallel continuations, the construction produced was annotated, distinguishing between non-elliptical VPs, VPE, *do it, do that* and *do so* constructions. **Table 4** provides an overview of the distribution of constructions in the data set.

Table 4: Frequency of constructions in the annotated data set.

Parallel VP	VPE	do it	do so	do that	other VP
313 (16.3%)	106 (5.5%)	36 (1.8%)	4 (0.2%)	4 (0.2%)	1453 (76.8%)

5.3.3 Embedding of the target VP

The next annotation layer tracked whether the parallel VP was embedded under another expression, such as a modal (13d), or even within an additional clause (14). Since one goal of Experiment 2 was to measure the predictability of the target VP in the context where it appears in Experiments 1 and 3, it was necessary to take into account the possible effects of material intervening between context and target VP.

(14)   a.   … had a telescope that allowed him to see it.
        b.   … looked extra hard and saw it.

For this reason, I classified (in principle) parallel VPs that were embedded in a further subordinate clause (14a) or in the second conjunct of a coordination (14b) as *other VP*. Both in (14a) and (14b), the parallel VP (*saw it*) appears somewhere in the continuation, but it is not

---

of (ia), they refer to the same action. In contrast, *its light* in (ib) was not categorized as a synonym, since the fact that the secretary saw the camera's light does not imply that she also saw the camera or realized that there was one, and the VP is hence not meaning-equivalent to the antecedent. Pronominalizations of the noun, as in (13), were also classified as lexically identical, since they imply coreference with the object DP contained in the antecedent.

(i)   a.   Joe was almost unbeatable, but in the end Sam defeated him.
        b.   The hidden camera was almost unnoticeable, but today the secretary saw its light blinking from afar.

conjoined with the first conjunct, as in the stimuli of Experiment 1. Therefore, continuations like (14) cannot contribute to a probability estimate which explains the omission of the VP immediately following the subject. Consequently, 112 VPs which would have otherwise been classified as parallel were categorized as *other VP* instead.[12]
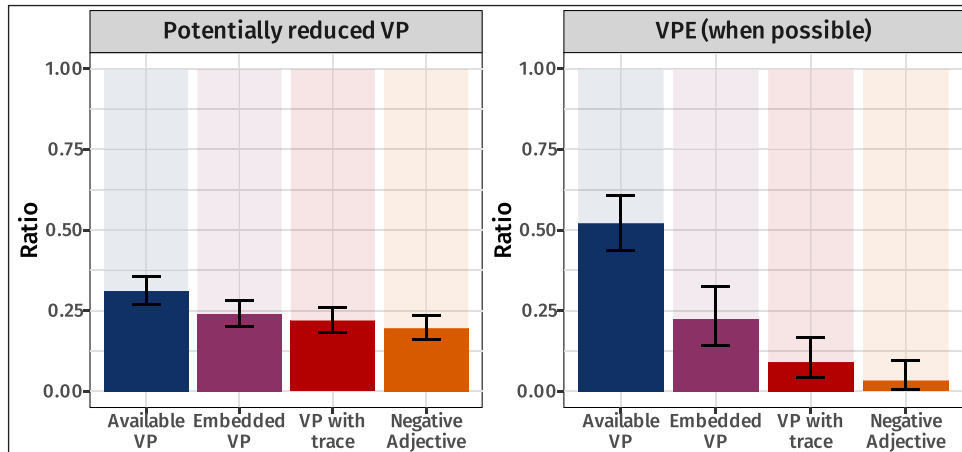

Figure 3: Ratio of parallel continuations that could be reduced by VPE (i.e. instances of VPE, parallel complete VPs, *do so*, *do it* and *do that* constructions) to the total responses (left) and ratio of VPE to the potentially reducible responses, i.e. parallel complete VPs and VPE (right). Error bars indicate 95% confidence intervals.

VPs and VPE embedded under only a modal or auxiliary were still classified as parallel, because VPE occurs also in the context of such verbs (15), so that the position in which the target VP would be omitted is, in principle, identical to the one where it is omitted under the auxiliary *did*. In the statistical analysis, whether or not the target was embedded was included as a predictor, to ensure that any effects on target VP predictability were not due to differences with respect to embedding.

(15)   None of the astronomers saw the comet, but John could.

5.4 Results

The analyses of the annotated production data addressed the two research questions introduced above: First, does the predictability of a potentially reduced VP gradually decrease across conditions, just as the perceived acceptability of VPE does? And, second, does the likelihood of these potentially reduced VPs actually being elided by participants also decrease accordingly?

The left panel of **Figure 3** provides an overview of the ratio of potentially reduced VPs to the total responses (after the exclusion of nonsense responses). The set of potentially reduced VPs

---

[12] The data in **Table 4** were aggregated after this classification.

comprises all complete VPs classified as parallel (see 5.3.1) and all instances of VPE. The right panel of **Figure 3** shows the ratio of VPE to this set of potentially reduced VPs. VPs that cannot be reduced are excluded here, in order to consider only utterances where VPE is an option. Both plots suggest that there is at least a gradual tendency in the expected direction, i.e. that both the potentially reduced VP is more predictable, and its omission more likely, in those conditions in which VPE is more acceptable.

The analyses reported in this section operate on the stricter of the two notions of identity discussed above, that is, lexical identity between the VP derived from the antecedent in the first conjunct and the one produced by the participant. For these responses, it is definitely the case that the VP can be omitted without a change in meaning. Applying the same set of analyses to the data set relying on the classification of synonym VPs as parallel yields very similar results.[13]

The data were analyzed with mixed effects logistic regressions (Bates et al., 2015) in R (R Core Team, 2022), which model the outcome of a binary dependent variable from a series of predictors. Again, likelihood ratio tests calculated with the anova function (R Core Team, 2022) were used to establish whether including a predictor in the model significantly improved model fit. I conducted two series of analyses to investigate the two research questions that motivated the experiment: first, whether the VP produced by participants is parallel to the one expected given the antecedent, and, second, whether this VP is reduced by VPE.

5.4.1 How predictable is a parallel VP?

I first investigated whether the predictability of a potentially elided VP gradually decreases from the available VP condition through the negative adjective condition, just like the acceptability ratings in Experiment 1. If this prediction were borne out, it would support the hypothesis that the acceptability of VPE increases with the predictability of the target VP.

The regression models predicted the outcome of a binary dependent variable PARALLELISM. In the main analysis, only lexically identical VPs, VPE and *do it*, *do that* and *do so* anaphora were

---

[13] I explored whether I obtain similar results when relying on the looser concept of identity, i.e. by categorizing VPs with synonym verbs and nouns (see 5.3), by running a series of analyses parallel to the one described in Sections 5.4.1 and 5.4.2. In the analysis of the target VP likelihood, only the contrast between available and embedded VPs is significant ($\chi^2 = 5.95$, $p < 0.05$), but not EVP v TRACE ($\chi^2 = 0.04$, $p > 0.8$) or TRACE v ADJ ($\chi^2 = 0.9$, $p > 0.3$). In the analysis predicting VPE, in addition to the effects found in the main analysis (see **Table 6**), there is a significant interaction between EMBEDDING and AVP v EVP ($\chi^2 = 14.82$, $p < 0.001$), which suggests that the preference for VPE when the target is embedded is weaker in the available VP condition than in the other ones. This is not directly predicted by the information-theoretic account, but it would be in line with it if potentially embedding verbs like modals increase the likelihood of a parallel VP and of VPE (as suggested by the EMBEDDING main effect). From an information-theoretic perspective, this would favor the usage of VPE in those conditions in which the target VP would be relatively unpredictable in the absence of the embedding verb.

classified as parallel, except when they were embedded in a different clause. For the reasons discussed above, in that case, they were also classified as non-parallel. The full model[14] contained fixed effects for the three Antecedent contrasts between the experimental conditions, which result from forward coding, as illustrated in **Table 1**, the scaled and centered Position of the trial in the experiment and all two-way interactions between these predictors. The model also contained by-subject and by-item random intercepts; it did not converge with a more complex random effects structure.

The final model (see **Table 5**) shows that the likelihood of a parallel VP is highest in the available VP condition ($\chi^2 = 23.43, p < 0.001$). While there is no significant difference between the embedded VP and VP with trace conditions ($\chi^2 = 0.5, p > 0.4$), a parallel VP is marginally less likely in the negative adjective condition than in the other ones ($\chi^2 = 3.04, p = 0.08$). Position had no significant effect and did also not interact with any of the Antecedent contrasts. Taken together, this suggests that the likelihood of a parallel VP decreases across conditions in the expected direction.

Table 5: Fixed effects in the final model predicting the likelihood of a parallel continuation.

Predictor	Estimate	SE	χ^2	p
Intercept	−1.74	0.24	28.5	<0.001
AVP v EVP	0.7	0.14	23.43	<0.001

5.4.2 How likely is VPE?

The second analysis investigated whether the likelihood of reducing a VP by ellipsis patterns with the likelihood of a parallel VP and the acceptability cline for VPE that the rating study revealed. For this analysis, I restricted the data set to complete parallel VPs and instances of VPE, excluding *do it*, *do that* and *do so* anaphora. Since these constructions densify utterances similarly to VPE, their usage might also be subject to information-theoretic constraints. Therefore, from the information-theoretic perspective, it would not be reasonable to treat them like complete VPs in the analysis. It was also not an option to pool them with the VPE data, since this would result in modeling the likelihood of any kind of reduction in that case, but this article is particularly concerned with VPE.[15,16]

---

[14] ParallelVP ~ (AVP v EVP + EVP v Trace + Trace v Adj) * Position + (1 | Subject) + (1 | Item).

[15] Furthermore, according to Hankamer and Sag (1976), *do it* is more tolerant of antecedent-target mismatches than VPE and *do so*, so treating all reductions as a single category might conceal differences between these constructions.

[16] A parallel analysis to the one reported in this section, pooling *do it*, *do that* and *do so* with VPE, yields a very similar pattern, probably due to the relatively low number of these constructions in the data set (see **Table 4** above).

Table 6: Fixed effects in the final model predicting the likelihood of VPE.

Predictor	Estimate	SE	χ^2	p
INTERCEPT	2.57	0.54	27.02	<0.001
AVP v EVP	–2.73	0.59	32.0	<0.001
EVP v TRACE	–1.3	0.6	8.08	<0.05
TRACE v ADJ	–1.6	0.87	3.88	<0.05
EMBEDDING	–1.22	0.55	8.28	<0.05

The full model predicting the likelihood of VPE contained the three forward-coded ANTECEDENT contrasts, the scaled and centered POSITION of the trial in the experiment, a binary predictor EMBEDDING, which tracked whether the target VP in the second conjunct was embedded (deviation coded as –0.5, 0.5), and all interactions (up to three-way) between these predictors, as well as by-subject and by-item random intercepts. Models with a more complex random effects structure did not converge. Including the EMBEDDING as a predictor allows for quantifying and isolating effects driven by the embedding construction. Otherwise, the procedure was identical to the previous analyses.[17]

The final model is summarized in **Table 6**. The significant main effects of all three ANTECEDENT contrasts show that the likelihood of VPE gradually decreases in the expected direction, i.e. from the available VP through the negative adjective condition. The main effect of EMBEDDING ($\chi^2 = 8.28, p < 0.05$) evidences that, overall, embedding the target VP increases the preference for ellipsis. There is no significant POSITION effect and none of the interactions between the predictors is significant.

5.5 Discussion

Experiment 2 had two main goals: First, to test whether the target VP differed in predictability across conditions, in line with the rating data, which would provide an information-theoretic explanation for the acceptability pattern. Second, to investigate whether the ratio of VPE produced by participants also varies as a function of the likelihood of the potential target VP.

As compared to the model summarized in **Table 6**, if other reductions are pooled with VPE, the main effects of TRACE v ADJ ($\chi^2 = 3.93, p = 0.08$) and EMBEDDING ($\chi^2 = 2.14, p > 0.1$) are not significant, whereas those of AVP v EVP ($\chi^2 = 12.05, p < 0.001$) and EVP v TRACE ($\chi^2 = 38.72, p < 0.001$) are. However, the former effect is still marginal, and the EMBEDDING effect is not theoretically relevant to the information-theoretic account.

[17] Ellipsis ~ (AVP v EVP + EVP v Trace + Trace v Adj) * Embedding * Position + (1 | Subject) + (1 | Item).

The first analysis of the production data showed that the target VP is, overall, more likely in those conditions where its omission is more acceptable. Even though not all of the contrasts between the four conditions were significant, the gradual decrease in the predictability of the target VP (quantified as the ratio of full or reduced parallel VPs) resembled the one found in the rating data. This supports the hypothesis that the gradual pattern in the rating data is actually driven by an overall tendency to reduce predictable expressions and to encode unpredictable ones more redundantly.

The second analysis showed that the production data are not only in line with the rating data, but that there is a similar gradual effect in the likelihood of subjects actually producing an instance of VPE, provided that this is possible. Participants also produced some instances of other VP reduction strategies (*do so*, *do it* and *do that*), but their low number did not allow for a systematic investigation of their distribution in comparison with VPE and complete VPs. The analysis also suggested that embedding the target VP under a modal increases the preference for VPE. This finding might also follow from an information-theoretic account, if the presence of an embedding expression like a modal or auxiliary increases the likelihood of the target VP.

The production data also show that participants produce antecedent-target mismatches in all conditions to some extent. From the perspective of the VP recycling hypothesis (Arregui et al., 2006) and the memory retrieval account (Parker, 2018), this is unexpected, because mismatches are analyzed as ungrammatical. These theories account for the observation that some presumably ungrammatical mismatches are perceived as relatively acceptable, but this does not explain why speakers would produce ungrammatical utterances on purpose. In contrast, under the information-theoretic account, there is no reason to assume that antecedent-target mismatches are ungrammatical; in fact, it implies that all utterances which are systematically produced by speakers are grammatical (leaving production errors aside). The observations that (i) participants systematically produced VPE mismatches, and that (ii) their ratio decreases with the predictability of the target are thus expected under the information-theoretic account.

An anonymous reviewer pointed out that the gradually decreasing ratio of VPE across conditions might be observed because the rarer mismatches violate more grammatical constraints. This could potentially reconcile the production of presumably ungrammatical structures with accounts which do not predict this. While it is possible that subjects randomly violate grammatical constraints and that the violation of more constraints in one utterance is less likely, the data speak against this. In the production data, there were no other violations of grammaticality beyond VPE, which suggests that the production of mismatches is more systematic than the result of random errors. Furthermore, the VP recycling hypothesis, in particular, is motivated by gradual differences between utterances which Arregui et al. (2006) view as categorically ungrammatical. If grammaticality were gradual, e.g. in an optimality-theoretic framework (Müller, 1999), where it decreases as a function of the number of constraints violated, there is no need to assume a processing component, because the grammar alone already predicts the observed pattern.

A further option is that the mismatch conditions do not gradually differ in acceptability, but that speakers deliberately produce ungrammatical ellipses because they are shorter, at least as long as the listener can still recover the intended meaning (Bergen & Goodman, 2015). If successful communication becomes less likely the more difficult it is to reconstruct (Arregui et al., 2006) or retrieve (Parker, 2018) the antecedent, a trade-off between the likelihood of success and the loss in efficiency would also predict a gradual pattern in the production data. However, Bergen and Goodman (2015) would predict other ungrammatical omissions, which also shorten the utterance and are easy to reconstruct (e.g. omissions of perfectly inferable articles or prepositions), but this is not attested in the data.

In sum, the production study confirms the central prediction of the information-theoretic account that the ratio of VPE varies as a function of the likelihood of the potential target of ellipsis. This supports an explanation of the pattern in the rating data observed by Arregui et al. (2006) and in Experiment 1 in terms of probability-driven reduction.

6. Experiment 3: Self-paced reading

Experiment 2 supports the prediction of the information-theoretic account that VPs which are more predictable from context are more often omitted. The production data are also in line with the information-theoretic explanation for the acceptability judgments, since VPE is also perceived as more acceptable in those conditions in which the target VP is more predictable. What the production study does not show directly is whether the probability and acceptability cline that they evidence is related to differences in processing effort. This is a crucial prediction of the information-theoretic account, according to which an unpredictable target VP results in VPE being degraded, due to the great effort required to process it, and – since speakers take into account the hearer's processing effort – also less likely to be produced.

I investigate this with a self-paced reading experiment which measures the effort required to process the ellipsis in the four conditions tested in Experiments 1 and 2, repeated in (16). Given the results of the production study, the information-theoretic account predicts that this effort gradually increases from (16a) through (16d), in line with the decrease in acceptability and production preference for VPE. In order to link processing effort and acceptability, the self-paced reading experiment was supplemented with a binary acceptability judgment task, for which I expected a similar pattern to the one observed in the acceptability rating Experiment 1.

6.1 Materials

The stimuli in (16) were, in principle, identical to those tested in Experiment 1. The only modification was the addition of a spillover region after the auxiliary *did*, which consists of a temporal clause introduced by *after* or a causal clause headed by *because*. Without a spillover region, the auxiliary *did* would be the last word of the utterance, so that there would be no

possibility of measuring the effort associated with its processing, which often is reflected in the reading times for words following a target word (Mitchell, 1984). Specifically for VPE, processing difficulty has been evidenced in the region past the auxiliary (Parker, 2022). The spillover region for (16) is given in (17).

(16)   a.   None of the astronomers saw the comet, but John did . . .   (Available VP)
           b.   Seeing the comet was nearly impossible, but John did . . .   (Embedded VP)
           c.   The comet was nearly impossible to see, but John did . . .   (VP with trace)
           d.   The comet was nearly unseeable, but John did . . .   (Negative adjective)

(17)   . . . because he had a special telescope.

6.2 Procedure

48 self-reported native speakers of American English, recruited on Prolific, participated in the web-based self-paced reading experiment, which was conducted using PCIbex (Zehr & Schwarz, 2018). None of them had participated in the previous experiments. Each participant was rewarded with £2.50. The participants read the sentences word by word by pressing the space bar in a centered self-paced reading paradigm. The stimuli were presented in black 20 pt Courier font on a white background. Before each trial, there was a separator allowing the participants to take a short break between trials, which they also could end by pressing the space bar. Each sentence was followed by a speeded binary acceptability judgment task. Subjects were asked to reply to the question "Is this sentence acceptable?" by pressing the Y(es) or the N(o) keys, respectively. This task had a 2000 ms timeout to ensure that subjects answered intuitively. In the instructions, participants were asked to answer quickly and notified about a time limit for answering. The participants were assigned to one of four lists in which the materials were distributed so that each subject saw each token set once and each condition equally often ($n = 4$). The 16 stimuli were mixed with 24 stimuli from an unrelated experiment on sluicing and 36 filler sentences. 24 of the filler sentences were coordinated structures, half of which involved subject gaps. The remaining 12 fillers were garden path sentences, half of which contained reduced relative clauses. The other half contained complementizer-less complement clauses. The stimuli were presented in individual pseudo-randomized order, ensuring that no trials of the same category (VPE, sluicing, filler) immediately followed each other. Four of the garden path and eight of the subject gap fillers were followed by an additional polar comprehension question presented after the acceptability judgment task, which was used as an attention check. The correct answer was "Yes" for half of the questions and "No" for the other half. 6 participants who responded correctly to less than 75% of the comprehension questions were excluded from further analysis.

6.3 Preprocessing of the reading time data

The reading time data analyses investigated the possible effects of increased effort of ellipsis processing in two regions: (i) on the auxiliary *did* itself, where readers become aware of VPE, and (ii) in the spillover region, where delayed effects of processing effort are expected to occur. I looked into effects at the onset of the spillover region, i.e. the first three words of the causal or temporal clause following the auxiliary.[18] Reading times faster than 90 ms and slower than 3000 ms (by word) were excluded from further analysis. This resulted in a loss of 0.45% of the total data points.

For the spillover region, the raw reading times for the first three words were first summed and then log-transformed using the natural logarithm. Regions with at least one missing data point, due to the removal of excessively fast and slow reading times described above, would have been excluded from further analysis, but none of the data points in the spillover region was affected. The log cumulated reading times were then residualized with a linear model (R Core Team, 2022) predicting the log cumulated reading time from the logarithmized position of the trial in the experiment and the cumulated length of the spillover region onset (measured in the number of characters). The residuals of this model were used as the dependent variable in the analysis of the spillover reading times.

The data for the auxiliary region were logarithmized using the natural logarithm. Since reading times were always measured on the same word (*did*), it was not necessary to factor out word length by residualizing response times.

6.4 Results

6.4.1 Self-paced reading data

Figure 4 provides an overview of the average reading times in the two critical regions (Auxiliary and Spillover onset). The data were analyzed with linear mixed effects regressions (Bates et al., 2015) in R (R Core Team, 2022), following a backwards model selection procedure as in the previous analyses described.

---

[18] Note also that in the case of the stimuli, after processing the auxiliary but not the spillover region, the sentence might still continue with a negation (ia) or *verum* focus construction (Höhle, 1988, 1992) like (ib), which contains both an auxiliary and an overt VP. If the parser's work consists of rejecting parses disconfirmed by the input (Hale, 2001) rather than adjusting a probability distribution over possible parses (Levy, 2008) no processing difficulties due to VPE are expected at the auxiliary at all.

(i)   None of the astronomers saw the comet . . .
   a.   . . . but John did not forget his telescope and saw it.
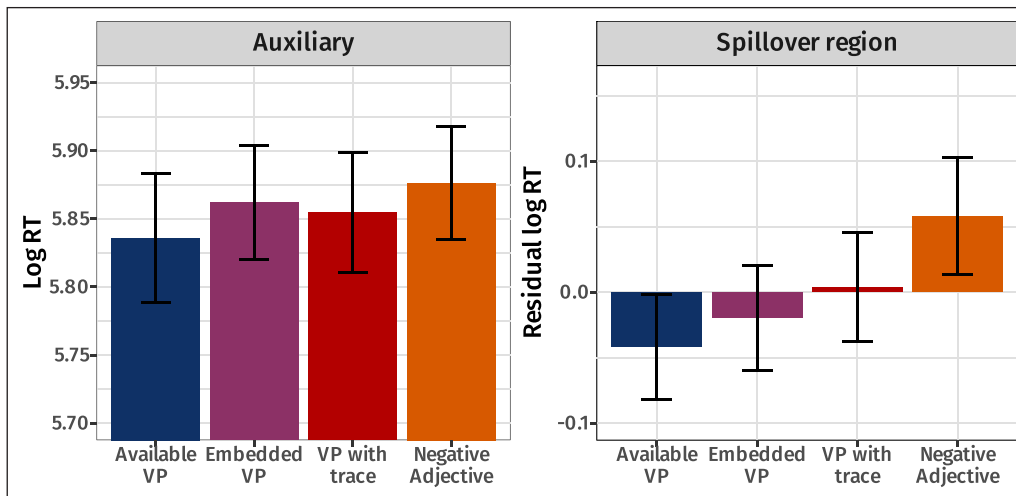   b.   . . . but John did see it.

Figure 4: Mean reading times by condition in the auxiliary region (left) and the onset of the spillover region (right). Error bars indicate 95% confidence intervals.

Table 7: Final model for the reading times on the auxiliary.

Predictor	Estimate	SE	χ^2	p
INTERCEPT	5.86	0.03	310.97	<0.001
POSITION	−0.06	0.01	73.59	<0.001

6.4.1.1 Auxiliary region

I first investigated possible effects of the experimental manipulation in the auxiliary region, in which participants might become aware of the omission of a VP. The full model[19] contained fixed effects for the three contrasts of ANTECEDENT, which was forward-coded as in the analyses of Experiments 1 and 2 (see **Table 1**), the scaled and centered POSITION of the trial in the experiment and their interactions. Furthermore, a by-subject random intercept was included. The model did not converge when a by-item intercept or higher-order random effects were included. The final model (see **Table 7**) contains only a theoretically uninteresting position effect, which reflects a general speed-up in the course of the experiment. None of the three ANTECEDENT contrasts was significant: AVP v EVP ($\chi^2 = 2.0$, $p > 0.1$), EVP v Trace ($\chi^2 = 0.09$, $p > 0.7$), Trace v Adj ($\chi^2 = 0.85$, $p > 0.3$).

---

[19] Log RT ~ (AVP v EVP + EVP v Trace + Trace v Adj) * Position + (1 | Subject).

6.4.1.2 Spillover region

The analysis for the spillover region used the residual log cumulated reading time for the first three words of the spillover region, which was computed as described in 6.3, as the dependent variable. Otherwise, the procedure was identical to the analysis of reading time on the auxiliary. Again, the full model contained the three ANTECEDENT contrasts, the scaled and centered POSITION of the trial in the experiment and their interactions as well as a by-subject random intercept.[20]

The final model is summarized in **Table 8**, and shows that VPE is processed significantly more slowly with a negative adjective antecedent than in the other conditions. Further simplifying this model was not possible, because removing either of the other two predictors basically pools two of the remaining conditions and results in the remaining additional contrast becoming significant: If AVP v EVP is removed, these two conditions taken together are significantly faster than the VP with trace condition ($\chi^2 = 4.91, p < 0.05$); if EVP v TRACE is removed, the available VP condition is read significantly faster than these two conditions taken together ($\chi^2 = 5.1, p < 0.05$).

Table 8: Fixed effects in the spillover region reading times model.

Predictor	Estimate	SE	χ^2	p
INTERCEPT	0.00	0.5	0	1.0
AVP v EVP	−0.02	0.02	1.58	>0.2
EVP v TRACE	−0.02	0.02	1.77	>0.1
TRACE v ADJ	−0.05	0.02	9.3	<0.01

Table 9: Dummy coding scheme for ANTECEDENT.

Predictor	Available VP	Embedded VP	VP with trace	Neg. adjective
AVP v TRACE	1	0	0	0
EVP v TRACE	0	1	0	0
ADJ v TRACE	0	0	0	1

---

[20] Residual log RT ~ (AVP v EVP + EVP v Trace + Trace v Adj) * Position + (1 | Subject).

The analysis using the forward coding scheme shows that the reading time data also exhibit a gradual pattern, and it confirms slower reading times for adjectival antecedents, but it does not allow one to determine which of the two other predictors should be removed. I therefore conducted a second analysis to test whether – within this gradual pattern – the available VP condition was read significantly faster than the VP with trace condition. For this purpose, I dummy-coded the ANTECEDENT predictor by setting the VP with trace condition as the reference level, and looked for significant pairwise differences between this predictor and the other ones (see **Table 9**). Except for the different coding scheme used, the full model in this analysis was identical to the one using the forward-coded predictor.

The final model (see **Table 10**) shows that besides the significant difference between the VP with trace and the negative adjective conditions ($\chi^2 = 18.22$, $p < 0.001$) there is also a significant difference between the available VP and the VP with trace conditions ($\chi^2 = 4.91$, $p < 0.05$).

Table 10: Fixed effects in the spillover region reading times model with Trace as baseline level.

Predictor	Estimate	SE	χ^2	p
INTERCEPT	−0.01	0.04	0.05	>0.8
AVP v TRACE	−0.03	0.02	4.91	<0.05
ADJ v TRACE	0.07	0.02	18.22	<0.001

6.4.2 Speeded acceptability judgment

Figure 5 provides an overview of the binary acceptability judgments and the mean response times to the judgment task across conditions. The judgments indicate that participants, overall, accepted all conditions except the negative adjective condition, which was rejected about half of the time. The response times suggest that positive responses were faster than negative responses, but there does not seem to be a systematic effect across conditions which is in line with the judgments.[21]

---

[21] A linear regression analysis (Bates et al., 2015) shows that positive judgments were provided faster than negative ones ($\chi^2 = 26.43$, $p < 0.001$) and that this was, in particular, the case for the available VP condition, as evidenced by an interaction between AVP v EVP and the polarity of the response ($\chi^2 = 4.87$, $p < 0.05$). Furthermore, a main effect of POSITION shows that response times decreased throughout the experiment, which indicates familiarization with the task.
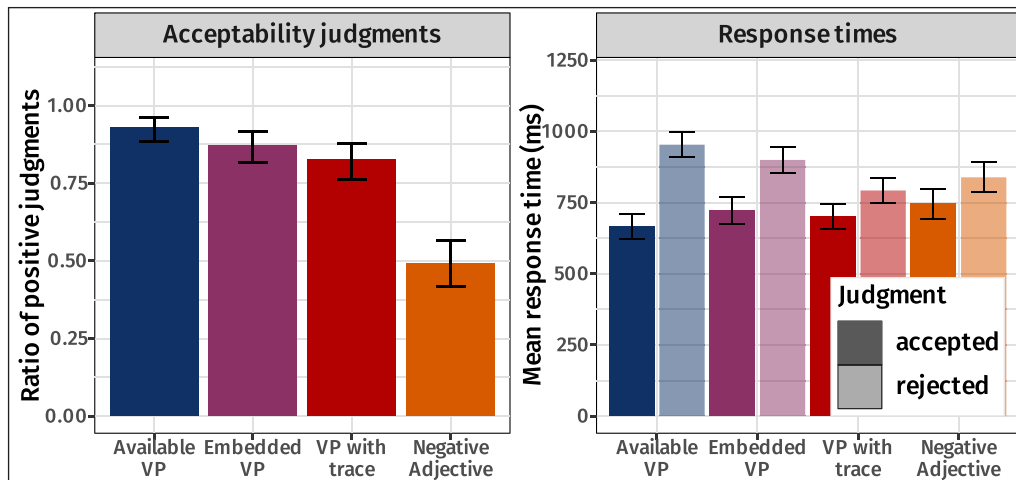
Figure 5: Ratio of positive judgments (left) and mean response times (right) across conditions in the binary acceptability judgment task. The error bars indicate 95% confidence intervals.

I used mixed effects logistic regressions to test whether the ratio of positive and negative judgments varies across conditions and whether it is related to the reading time on the spillover region, which indicates gradually increasing processing difficulty. Since the information-theoretic account predicts that VPE is degraded when it is hard to process, I expected the ratio of negative judgments to increase as a function of reading time. In addition to the categorical differences between conditions, including the actual reading time for individual stimuli might indicate differences between stimuli and individual participants.

The analysis investigated the effects of the ANTECEDENT (forward-coded), the scaled and centered POSITION of the trial in the experiment, the residual log READING TIME for the spillover region (which was the dependent variable in the reading time analysis, continuous) and their interactions with mixed effects logistic regressions predicting the outcome of a binary dependent variable JUDGMENT. Timed-out trials ($n = 22$) were excluded from the data set for this analysis, because their overall number was relatively small and the analysis with logistic regressions requires a binary dependent variable (in this case, positive and negative judgments). The full model contained fixed effects for the three ANTECEDENT contrasts and POSITION and READING TIME, their interactions, and by-subject and by-item random intercepts. Models with a more complex random effects structure did not converge. Otherwise, the procedure was identical to the logistic regression analyses of the production data.[22]

---

[22] Judgment ~ (AVP v EVP + EVP v Trace + Trace v Adj) * Position * Reading Time + (1 | Subject) + (1 | Item).

The final model is summarized in **Table 11**. The main effect of Trace v Adj confirms that VPE was significantly more likely to be rejected with an adjectival antecedent than in any of the other conditions ($\chi^2 = 83.54, p < 0.001$). The effect of AVP v EVP suggests that VPE with available antecedents was less often rejected than in the other conditions ($\chi^2 = 8.3, p < 0.01$). The EVP v Trace contrast comparing the medial levels was not significant ($\chi^2 = 1.79, p > 0.1$). This supports a gradual acceptability cline across the four conditions, which has been found in the previous experiments. The significant effect of Reading Time shows that VPE is perceived as more acceptable when it is easier to process. This effect is driven by individual differences between subjects and differences between token sets, since the gradually increasing processing effort between conditions, which has been evidenced by the reading time data analysis, is already explained by the Antecedent effects included in the model.

Table 11: Final model for the binary acceptability judgment data.

Predictor	Estimate	SE	χ^2	p
Intercept	1.89	0.25	32.76	<0.001
AVP v EVP	0.96	0.35	8.3	<0.01
Trace v Adj	0.96	0.25	83.54	<0.001
Reading Time	−1.64	0.49	11.56	<0.001

6.5 Discussion

Experiment 3 used a combined self-paced reading and binary acceptability judgment paradigm to investigate whether VPE is harder to process in those conditions in which the omitted VP is less predictable and VPE more strongly degraded. The results confirm a crucial prediction of the information-theoretic account of antecedent-target mismatches, given the differences in VP predictability observed in the production study: Speakers avoid utterances which induce excessively high peaks of processing effort, and hearers perceive these as more heavily degraded. The self-paced reading study was supplemented with a speeded binary acceptability rating task in order to relate processing effort to the perceived acceptability of VPE.

6.5.1 Reading time data

The reading time data reveal a gradual increase in processing difficulty in the spillover region, which is aligned with a decrease in likelihood of the target VP and, consequently, of the acceptability of VPE. The pattern seems to be more gradual between the three more acceptable conditions,

so that only a comparison between the available VP and the VP with trace conditions reaches significance in the main analysis.[23] However, overall, it exhibits the same gradual tendency as the data from the previous two experiments, and it particularly resembles the pattern observed in the binary judgment task, where the largest difference between subsequent conditions is also found between the VP trace condition and the negative adjective condition. In the auxiliary region, there appears to be a tendency in the same direction, but the contrasts between the conditions are not significant. This might be due to the auxiliary being only a three-character word which is easily recognized and skipped over by participants, whereas ellipsis processing takes place while reading the subsequent words. Furthermore, at this point in the utterance, participants do not know for sure that the sentence is elliptical, since it might also continue as a nonelliptical sentence containing an auxiliary, e.g. the in case of negation or *verum* focus.

6.5.2 Binary judgments

The binary acceptability judgments follow a similar pattern as the previous experiments, that is, a decrease in acceptability from the available VP through the negative adjective condition. In addition to the significant contrasts between conditions, which are related to increasing processing effort (as the reading time analysis showed), individual reading times additionally predict the acceptability of individual trials: Subjects who read ellipses faster are also more likely to judge an instance of VPE as grammatical.

In comparison to the Likert scale data from Experiment 1, the negative adjective condition was particularly degraded and the differences between the other conditions less pronounced. This might be due to the different paradigm, which does not allow one to gradually judge a single trial. It also differs from the pattern reported by Arregui et al. (2006, p. 235), who find a decrease from 82.8% (available VP), in steps of about 20%, through 17.1% for adjectival antecedents in a binary acceptability judgment task. This might be due to two differences in the methodology: First, I put participants under time pressure, so they had less time to evaluate acceptability than in the study by Arregui et al. (2006). Second, in their experiment, the stimulus was displayed during the judgment task, whereas it was not in mine, so subjects could not re-read the stimulus before providing their judgment. Being able to read the stimulus carefully for an unlimited time might result in more deliberate judgments.

---

[23] An anonymous reviewer suggested that the nonsignificance of some of the effects might result from the study being underpowered. However, there are significant differences between conditions in the spillover region, and the greater difference between the negative adjective condition and the other three conditions that reading times reveal is in line with the binary acceptability judgments. It might still be interesting to replicate the study with more subjects and materials in the future.

6.5.3 Summary

Experiment 3 confirms the prediction of the information-theoretic account that VPE is more acceptable when it is easier to process. This prediction is shared with the VP recycling hypothesis and the memory retrieval account. The analysis of the binary acceptability judgments showed that the effect of reading times holds not only between the experimental conditions, but beyond the variation explained by the contrasts between conditions. From an information-theoretic perspective, this might be due to differences in VP predictability varying between token sets and conditions or individual differences between participants. Systematic variation between subjects is also expected under the VP recycling hypothesis and the memory retrieval account, if syntactic derivation or memory retrieval requires more processing effort for some participants in general.

7. Item level analyses

The results of the three experiments are in line with an information-theoretic explanation for the acceptability cline observed by Arregui et al. (2006): The more predictable the target VP is in a condition, the faster VPE is processed, which results in higher acceptability for VPE. However, since the information-theoretic account derives the other measures from the predictability of the target VP, it makes actually more fine-grained predictions at the level of individual stimuli: If the VP is more likely in a particular stimulus in a specific condition than in another stimulus in the same condition (for whatever reason), the acceptability and processing effort are expected to vary accordingly.[24] This distinguishes the information-theoretic account from the VP recycling hypothesis (Arregui et al., 2006) and the memory retrieval account (Parker, 2018). While all three theories share the predictions concerning the acceptability rating and reading time data, the latter two accounts do not predict systematic differences between stimuli: They derive processing effort from the syntactic similarity between antecedent and target, which is systematically varied between, but not within, conditions.

The analysis of the judgment data collected in Experiment 3 already showed an effect of processing effort on acceptability beyond the experimental conditions. In what follows, I investigate whether the relationship between the predictability of the potential target of ellipsis, the acceptability of VPE and processing effort also holds at the level of individual stimuli. All analyses rely on the aggregated dependent variables by token set and condition, and effects between measures are compared using linear regressions in R (R Core Team, 2022), which investigate whether the dependent measures collected in the experiments predict each other beyond the effect of the experimental condition. The analyses presented in this section are somewhat exploratory, since the materials were not designed to look into item effects, for instance, by systematically manipulating VP predictability within each of the four ANTECEDENT conditions.

---

[24] I thank an anonymous reviewer for suggesting that I look into possible effects at the item level.

Table 12: Fixed effects in the final model predicting acceptability judgments from predictability.

Predictor	Estimate	SE	*F*	p
INTERCEPT	3.78	0.2	1130.1	<0.001
PARALLEL CONTINUATION	2.0	0.72	7.69	<0.01
AVP v EVP	0.79	0.26	9.18	<0.01
EVP v TRACE	0.57	0.26	4.97	<0.05
TRACE v ADJ	1.44	0.26	32.0	<0.001

7.1 Does VP predictability predict acceptability?

The first analysis investigates the information-theoretic account's prediction that the omission of more predictable expressions is perceived as more acceptable. The production data from Experiment 2 already showed that the gradual acceptability pattern between conditions can be attributed to a difference in likelihood, but not whether acceptability differences between stimuli can, too. To investigate this, I aggregated the acceptability ratings from Experiment 1 and the ratios of parallel continuations from Experiment 2 by item and condition and performed a linear regression analysis (R Core Team, 2022) to test whether the production data predict ratings.

The full model predicted the mean ACCEPTABILITY (numeric, aggregated by item and condition) from the likelihood of a PARALLEL CONTINUATION (numeric), the three ANTECEDENT contrasts (forward-coded) and all two-way interactions.[25] The ANTECEDENT contrasts were included to factor out the differences between conditions found in Experiment 1 and distinguish additional effects of PARALLEL CONTINUATION from the overall pattern. The final model (see **Table 12**), which was obtained by backward model selection, as in previous analyses, contains significant main effects of all three ANTECEDENT contrasts, which replicate the pattern found in the rating study. The additional PARALLEL CONTINUATION effect ($F = 7.69$, $p < 0.01$) shows that the likelihood of the target additionally increases the acceptability of VPE. This indicates that VPE is more acceptable when the target is likely, not only between conditions, but also within a single condition.

7.2 Does reading time predict acceptability?

The analysis of the binary judgment data already showed that reading time has an effect on acceptability beyond the gradual difference between conditions in the spillover region, which was

---

[25] Acceptability ~ (AVP v EVP + EVP v Trace + Trace v Adj) * Parallel continuation.

revealed by the reading time data analysis. The analysis in this subsection investigates whether the same holds for the Likert scale acceptability ratings for VPE collected in Experiment 1.

The full model predicted the mean acceptability by token set and condition from the three forward-coded ANTECEDENT contrasts, the aggregated residual log READING TIME in the spillover region for each token set and condition and all interactions between the predictors.[26] The final model (see **Table 13**) shows that acceptability gradually decreases across conditions (all contrasts are significant). Additionally, a main effect of READING TIME ($F = 4.65$, $p < 0.05$) reveals that sentences which are read more slowly are perceived as less acceptable. This shows that not only the binary judgments but also the Likert scale ratings for VPE improve when processing the ellipsis is easier.

Table 13: Fixed effects in the final model predicting acceptability time from reading time.

Predictor	Estimate	SE	F	p
INTERCEPT	4.27	0.09	2149.8	<0.001
READING TIME	1.96	0.91	4.65	<0.05
AVP v EVP	0.97	0.26	13.9	<0.001
EVP v TRACE	0.66	0.26	6.3	<0.05
TRACE v ADJ	1.59	0.26	36.04	<0.001

7.3 Does VP predictability predict reading time?

A third analysis investigated whether the predictability of the target VP has an effect on reading time in the spillover region. It is a well-established finding that cloze probability predicts reading times (see e.g. Smith & Levy, 2011) and my Experiments 2 and 3 support such an effect at the condition level, but not for individual stimuli.

The analysis started from a full model predicting the mean READING TIME (numeric, aggregated by item and condition) from the likelihood of a PARALLEL CONTINUATION (numeric), the forward-coded ANTECEDENT contrasts and all two-way interactions.[27] The final model is summarized in **Table 14**. Like in the analyses of the reading time data in 6.4.1, VPE is processed significantly more slowly in the negative adjective condition. The other contrasts between conditions are not significant in this analysis (for AVP v EVP, $F = 3.04$, $p = 0.09$, for EVP v TRACE, $F = 0.67$, $p > 0.4$). The significant effect of PARALLEL CONTINUATION ($F = 4.65$, $p < 0.05$) suggests that in

---

[26] Acceptability ~ (AVP v EVP + EVP v Trace + Trace v Adj) * Reading Time.

[27] Reading Time ~ (AVP v EVP + EVP v Trace + Trace v Adj) * Parallel continuation.

addition to the Antecedent effects, which resemble the pattern found in Experiment 3, there is an item level effect suggesting that VPE is processed more slowly when the target is predictable. This is unexpected under the information-theoretic account, which predicts the opposite.

Table 14: Fixed effects in the final model predicting reading time from VP predictability.

Predictor	Estimate	SE	*F*	p
Intercept	–0.05	0.03	3.62	$= 0.06$
Parallel Continuation	0.21	0.1	4.65	< 0.05
Trace v Adj	–0.09	0.03	9.14	< 0.01

7.4 Discussion

The analyses in this section investigated to what extent the acceptability, predictability and processing effort measures collected in the experiment are related not only in the gradual pattern across conditions, but also at the level of individual stimuli, in the way the information-theoretic account predicts. This was investigated with regression analyses testing for effects of continuous reading time and predictability predictors beyond the contrasts between the four conditions, which have been established in the analyses of the experiments in Sections 4–6.

The predictions of the information-theoretic account are supported by two of the three analyses: First, VPE is perceived as more acceptable when the potential target VP is more predictable. From the information-theoretic perspective, this evidences a tendency to reduce predictable VPs and to use more redundant encodings for unpredictable VPs. Second, acceptability also decreases as a function of processing effort, indexed by reading times. This holds both for the binary judgments collected in Experiment 3 and the Likert scale data from Experiment 1. The expectation that a greater VP predictability similarly leads to reduced processing effort was not confirmed. Note that the inverted effect of VP predictability on reading times does not indicate that there is no effect at all of VP predictability on reading times, but that in this analysis, there is no additional effect of VP predictability beyond the differences between conditions. Since the materials were not designed to perform the analyses reported in this section, future research might look further into this by employing of stimuli that systematically vary VP predictability beyond the Antecedent condition.[28]

---

[28] A possibility might be to add further context to the stimuli that manipulates the likelihood of the target VP through world knowledge, similarly to an experiment by Schäfer et al. (2021). For instance, world knowledge might make the target VP *saw the comet* more likely in context (ia) than in (ib), without altering the syntactic structure.

    (i)   None of the astronomers saw the comet, but John did.
        a.   It was a clear night.
        b.   It was a cloudy night.

Except for the unexpected effect of VP predictability on reading times, these results are in line with the information-theoretic account. The effect of reading times on acceptability might also be expected under the VP recycling hypothesis (Arregui et al., 2006) and the memory retrieval account (Parker, 2018), because they derive acceptability from processing effort: Whatever makes certain stimuli difficult to process in addition to the reconstruction or memory retrieval effort will also result in degraded acceptability. However, neither of these accounts predicts that VP predictability affects acceptability. The predictability effect does not falsify their predictions, but they do not provide an explanation for it, while the information-theoretic account does.

8. General discussion

In this article, I presented three experiments testing whether the gradual acceptability cline of antecedent-target mismatches under VPE reported by Arregui et al. (2006) can be explained by an information-theoretic account of language processing and production. Such an account predicts that the likelihood of a potential target of ellipsis (in this case, a VP) actually being omitted increases as a function of its predictability in context: Omitting predictable material increases the efficiency of communication, whereas selecting more explicit forms for unpredictable expressions avoids overutilizing the hearer's processing resources.

The experiments reported in this article support the central predictions of the information-theoretic account: First, VPE is perceived as more acceptable when the target VP is more predictable in context. Second, the VP is also more likely to be actually omitted in these conditions in the production data, so that production preferences are aligned with the perceived well-formedness of utterances. Third, the reading time data show that VPE is harder to process when the target VP is less predictable (and VPE less acceptable). Taken together, these findings suggest that the likelihood of the potential target of VPE affects the processing effort and acceptability of VPE as well as the speakers' production choices, which adapt the utterance to the hearer's processing effort. This provides an information-theoretic explanation of the acceptability cline observed by Arregui et al. (2006) across conditions. As the analyses in Section 7 showed, most of the predictions of the information-theoretic account are also supported by additional effects at the level of individual token sets, except for the expected effect of VP predictability on the effort required to process the VP.

8.1 Implications for other theories of VPE mismatches

The results of the experiments are also partially in line with some of the other accounts of VPE mismatches discussed in Section 2. In what follows, I discuss the implications of my findings with respect to the theories discussed there.

8.1.1 Syntactic approaches

Purely syntactic accounts of VPE mismatches, like Merchant (2013) or Johnson (2001), might explain why some mismatches are (un)grammatical, but not why we observe gradual usage preferences among these mismatches. However, the specific proposals are not designed to account for the specific types of mismatches tested in this article. Syntactic identity conditions might be specified to fit the data, e.g. in an optimality-theoretic framework based on violable constraints, but this would require independent evidence of their effect.

8.1.2 Parsing heuristics

The processing-based account of mismatch acceptability of C. S. Kim et al. (2011) is similar to the information-theoretic account with respect to the assumption that the expectations of the parser determine the processing effort and acceptability of VPE.[29] However, my data are not in line with their prediction of a preference for realizing the complement to the right of the transitive verbal head, which follows from their canonical word order constraint. In the materials tested in my experiments, this distinguishes the available and embedded VP conditions (*saw the comet/seeing the comet*) from the VP trace and negative adjective conditions (*the comet was nearly impossible to see/the comet was nearly unseeable*). Since the complement *the comet* appears right of the head *see*) only in the former two conditions, C. S. Kim et al. (2011) predict these to be more acceptable. Even though this is the case in most of my experiments, the differences between the medial conditions (EVP and VP trace) were much less pronounced than the differences between these and the ones at the ends of the scale (AVP and neg. Adjective). Since the processing heuristics proposed in C. S. Kim et al. (2011) do not explain the more pronounced contrasts between the two extreme and the medial conditions, they fail to account for the complete pattern observed.

8.1.3 Information structure

Kertz (2013) argues that VPE mismatches are licensed if a contrastive topic relation holds between the clause containing the antecedent and the one containing the target of VPE. In the materials, repeated here as (18), this is only the case in the available VP condition (18a), which is acceptable anyway, because it does not involve a mismatch. Therefore, the information-structural account does not explain the gradual contrasts between the three mismatch conditions in (18b–d). This does not speak against Kertz' account, but it shows that information structure does not suffice to explain the pattern observed in my experiments.

---

[29] The assumption of C. S. Kim et al. (2011) that the parser is guided by specific heuristics is not shared by information-theoretic accounts, which presuppose a parallel parser (Hale, 2001; Levy, 2008). This does not rule out the possibility that grammatical rules restrict the set of possible parses or that the heuristics actually reflect differences in probability between constructions.

(18) a. None of the astronomers saw the comet, but John did. (Available VP)
      b. Seeing the comet was nearly impossible, but John did. (Embedded VP)
      c. The comet was nearly impossible to see, but John did. (VP with trace)
      d. The comet was nearly unseeable, but John did. (Negative adjective)

8.1.4 Discourse and extralinguistic context

Since I tested individual utterances and a single discourse connective (*but,* which does not encode parallel discourse relations, according to Kehler (2000)), my data does not allow for conclusions concerning theories of VPE mismatches which take into account discourse or extralinguistic context (Geiger & Xiang, 2021). The materials did also not contain particular implicature or QuD triggers that could license VPE mismatches in a systematic fashion in some conditions, as proposed by Grant et al. (2012) and Miller and Hemforth (2014), respectively.

8.1.5 VP recycling hypothesis

The data are partially in line with the VP recycling hypothesis of Arregui et al. (2006), on whose stimuli I based my experiments, since their predictions concerning reading times and acceptability are identical to those of the information-theoretic account. From their perspective, the reading time data would be interpreted as showing that processing ungrammatical utterances takes more time, due to the reconstruction process involved. However, they do not explain the systematic production of VPE mismatches, which I observed in the production experiment, and the effects at the item level, which cannot be traced back to a difference in the steps in the syntactic derivation required to construct a matching antecedent.[30] Even though most of my results are in line with the VP recycling hypothesis, the information-theoretic account explains the data with independently evidenced processing mechanisms, whereas Arregui et al. (2006) require a parsing mechanism which is needed only to process VPE mismatches.

8.1.6 Memory retrieval

Since the memory retrieval account of Parker (2018) predicts the same acceptability pattern as Arregui et al. (2006) and also attributes degraded acceptability to greater processing effort, his predictions are also in line with the acceptability rating and self-paced reading data. Similarly to the approach taken in this article, Parker explains the empirically observed acceptability cline with an independently required processing mechanism, memory retrieval, which has independently been shown to modulate processing effort and acceptability. Like the VP recycling hypothesis, Parker (2018) analyzes mismatches as ungrammatical, and while he explains why

---

[30] Under the assumption of audience design, the gradual cline in the production data could possibly be explained by a tendency for speakers not to produce utterances which are too difficult to process.

some mismatches might be acceptable, he does not predict that speakers will produce those mismatches. Since his empirical predictions are aligned with Arregui et al. (2006), he also does not predict the systematic production of mismatches and the effects at the item level.

8.1.7 Implications for other accounts – Summary

In sum, the data are relatively consistent with the VP recycling hypothesis and the memory retrieval account, except for the systematic production of mismatches and the item level effects. The information-theoretic account predicts that the ratio of mismatches produced increases as a function of predictability, while the other two approaches analyze mismatches as ungrammatical (even though they might predict a similar effect under the assumption of audience design). Furthermore, the analysis in 7.1 showed an effect of VP predictability at the item level on VPE acceptability, which is expected only under the information-theoretic perspective. The other accounts of mismatches either cannot explain the gradual pattern found in the experiments, predict different effects, or focus on effects of discourse or extralinguistic context which were not manipulated in my experiments.

8.2 Conclusion

The experiments and analyses presented support the main predictions of an information-theoretic account of antecedent-target mismatches under VPE: VPE is more acceptable when the potential target VP is more predictable and processing the ellipsis requires less processing effort. This provides an account of the acceptability cline reported by Arregui et al. (2006) in terms of an independently evidenced processing mechanism, which has also been shown to account for other omission and reduction phenomena.

Data availability statement

Ethics and consent

Acknowledgements

Competing interests

The author has no competing interests to declare.

ORCiD IDs

Robin Lemke: 0000-0003-2964-7396

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

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

Aylett, M., & Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*(1), 31–56. https://doi.org/10.1177/00238309040470010201

Barros, M. (2014). *Sluicing and identity in ellipsis* [Doctoral dissertation, Rutgers University].

Barros, M., & Kotek, H. (2019). Ellipsis licensing and redundancy reduction: A focus-based approach. *Glossa: A Journal of General Linguistics*, *4*(1). https://doi.org/10.5334/gjgl.811

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bergen, L., & Goodman, N. (2015). The strategic use of noise in pragmatic reasoning. *Topics in Cognitive Science*, *7*(2), 336–350. https://doi.org/10.1111/tops.12144

Christensen, R. H. B. (2022). *Ordinal—Regression Models for Ordinal Data*. R package version 2022.11-16, https://CRAN.R-project.org/package=ordinal.

Chung, S. (2006). Sluicing and the lexicon: The point of no return. *Annual Meeting of the Berkeley Linguistics Society*, *31*, 73–91. https://doi.org/10.3765/bls.v31i1.896

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

Cuskley, C., Bailes, R., & Wallenberg, J. (2021). Noise resistance in communication: Quantifying uniformity and optimality. *Cognition*, *214*, 104754. https://doi.org/10.1016/j.cognition.2021.104754

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*, 193–210. https://doi.org/10.1016/j.cognition.2008.07.008

Elbourne, P. (2008). Ellipsis sites as definite descriptions. *Linguistic Inquiry*, *39*(2), 191–220. https://doi.org/10.1162/ling.2008.39.2.191

Fenk, A., & Fenk, G. (1980). Konstanz im Kurzzeitgedächtnis – Konstanz im sprachlichen Informationsfluß. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *27*(3), 400–414.

Frank, A. F., & Jaeger, T. F. (2008). Speaking rationally: Uniform Information Density as an optimal strategy for language production. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *30*, 939–944.

Geiger, J., & Xiang, M. (2021). At the syntax-discourse interface: Verb phrase ellipsis interpretation in context. *Language*, *97*(1), e89–e110. https://doi.org/10.1353/lan.2021.0010

Ginzburg, J., & Sag, I. A. (2000). *Interrogative investigations: The form, meaning, and use of English interrogatives*. CSLI.

Grant, M., Clifton, C., & Frazier, L. (2012). The role of Non-Actuality Implicatures in processing elided constituents. *Journal of Memory and Language*, *66*(1), 326–343. https://doi.org/10.1016/j.jml.2011.09.003

Hale, J. T. (2001). A probabilistic Earley parser as a psycholinguistic model. *Proceedings of NAACL (Vol. 2)*, 159–166. https://doi.org/10.3115/1073336.1073357

Hankamer, J., & Sag, I. A. (1976). Deep and surface anaphora. *Linguistic Inquiry*, *7*(3), 391–428.

Hardt, D. (1993). *Verb phrase ellipsis: Form, meaning, and processing* [Doctoral dissertation, University of Pennsylvania].

Hardt, D. (1999). Dynamic interpretation of verb phrase ellipsis. *Linguistics and Philosophy*, *22*(2), 185–219. https://doi.org/10.1023/A:1005427813846

Hardt, D., & Romero, M. (2014). Ellipsis and the structure of discourse. *Journal of Semantics*, *21*, 375–414. https://doi.org/10.1093/jos/21.4.375

Hendriks, P. (2004). Coherence relations, ellipsis and contrastive topics. *Journal of Semantics*, *21*(2), 133–153. https://doi.org/10.1093/jos/21.2.133

Hofmeister, P., Casasanto, L. S., & Sag, I. A. (2013). Islands in the grammar? Standards of evidence. In J. Sprouse & N. Hornstein (Eds.), *Experimental syntax and island effects* (pp. 42–63). Cambridge University Press. https://doi.org/10.1017/CBO9781139035309.004

Höhle, T. N. (1988). Vorwort und Nachwort zu *Verum-Fokus*. *Sprache und Pragmatik*, *5*, 1–7.

Höhle, T. N. (1992). Über Verum-Fokus im Deutschen. In J. Jacobs (Ed.), *Informationsstruktur und Grammatik* (pp. 112–141). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-663-12176-3_5

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62. https://doi.org/10.1016/j.cogpsych.2010.02.002

Johnson, K. (2001). What VP ellipsis can do, and what it can't, but not why. In M. Baltin & C. Collins (Eds.), *The Handbook of Contemporary Syntactic Theory* (pp. 439–479). Blackwell. https://doi.org/10.1002/9780470756416.ch14

Kehler, A. (2000). Coherence and the resolution of ellipsis. *Linguistics and Philosophy*, *23*, 533–575. https://doi.org/10.1023/A:1005677819813

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. CSLI.

Kennedy, C. (2003). Ellipsis and syntactic representation. In K. Schwabe & S. Winkler (Eds.), *Linguistik Aktuell/Linguistics Today* (pp. 29–53, Vol. 61). John Benjamins. https://doi.org/10.1075/la.61.03ken

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

Kim, C. S., Kobele, G. M., Runner, J. T., & Hale, J. T. (2011). The acceptability cline in VP ellipsis. *Syntax*, *14*(4), 318–354. https://doi.org/10.1111/j.1467-9612.2011.00160.x

Kim, C. S., & Runner, J. T. (2011). Discourse structure and syntactic parallelism in VP ellipsis. *UMass Occasional Papers in Linguistics*, *38*.

Kim, N., Carlson, K., Dickey, M., & Yoshida, M. (2020). Processing gapping: Parallelism and grammatical constraints. *Quarterly Journal of Experimental Psychology*, *73*(5), 781–798. https://doi.org/10.1177/1747021820903461

Kravtchenko, E. (2014). Predictability and syntactic production: Evidence from subject omission in Russian. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*, 785–790.

Kurumada, C., & Jaeger, T. F. (2015). Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language*, *83*, 152–178. https://doi.org/10.1016/j.jml.2015.03.003

Lemke, R. (2021). *Experimental investigations on the syntax and usage of fragments*. Language Science Press. https://doi.org/10.5281/zenodo.5596236

Lemke, R., Schäfer, L., & Reich, I. (2022). Can identity conditions on ellipsis be explained by processing principles? In R. Hörnig, S. von Wietersheim, A. Konietzko, & S. Featherston (Eds.), *Linguistic Evidence 2020 proceedings – Linguistic theory enriched by experimental data* (pp. 541–561). Universität Tübingen. https://doi.org/10.15496/publikation-75887

Levy, R. P. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Levy, R. P., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing* (pp. 849–856, Vol. 19). MIT Press. https://doi.org/10.7551/mitpress/7503.003.0111

LimeSurvey GmbH. (2012). *LimeSurvey: An open source survey tool*.

Mahowald, K., Dautriche, I., Gibson, E., & Piantadosi, S. T. (2018). Word forms are structured for efficient use. *Cognitive Science*, 1–19. https://doi.org/10.1111/cogs.12689

Merchant, J. (2004). Fragments and ellipsis. *Linguistics and Philosophy, 27*(6), 661–738. https://doi.org/10.1007/s10988-005-7378-3

Merchant, J. (2008). Variable island repair under ellipsis. In K. Johnson (Ed.), *Topics in ellipsis* (pp. 132–153). Cambridge University Press. https://doi.org/10.1017/CBO9780511487033.006

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

Miller, P., & Hemforth, B. (2014). VP ellipsis beyond syntactic identity: The case of nominal antecedents. https://doi.org/10.13140/2.1.4713.2488

Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods of investigating immediate processes in reading. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 69–90). Erlbaum. https://doi.org/10.4324/9780429505379-4

Müller, G. (1999). Optimality, markedness, and word order in German. *Linguistics, 37*, 777–818. https://doi.org/10.1515/ling.37.5.777

Norcliffe, E., & Jaeger, T. F. (2016). Predicting head-marking variability in Yucatec Maya relative clause production. *Language and Cognition, 8*(2), 167–205. https://doi.org/10.1017/langcog.2014.39

Parker, D. (2018). A memory-based explanation of antecedent-ellipsis mismatches: New insights from computational modeling. *Glossa: A Journal of General Linguistics, 3*(1), 129. https://doi.org/10.5334/gjgl.621

Parker, D. (2022). Ellipsis interference revisited: New evidence for feature markedness effects in retrieval. *Journal of Memory and Language, 124*, 104314. https://doi.org/10.1016/j.jml.2022.104314

R Core Team. (2022). *R: A language and environment for statistical computing*.

Sag, I. A. (1976). *Deletion and logical form* [Doctoral dissertation, MIT].

Schäfer, L., Lemke, R., Drenhaus, H., & Reich, I. (2021). The role of UID for the usage of Verb Phrase Ellipsis: Psycholinguistic evidence from length and context effects. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.661087

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal, 27*(4), 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x

Smith, N. J., & Levy, R. P. (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing, 1637–1642.

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

Tily, H., & Piantadosi, S. T. (2009). Refer efficiently: Use less informative expressions for more predictable meanings. *Proceedings of the Workshop on the Production of Referring Expressions: Bridging the Gap between Computational and Empirical Approaches to Reference.*

van Craenenbroeck, J. (2012). *Ellipsis, identity, and accommodation.* Ms. KU Leuven HU Brussel.

Williams, E. (1977). Discourse and logical form. *Linguistic Inquiry, 8*, 101–139.

Yoshida, M., Dickey, M. W., & Sturt, P. (2013). Predictive processing of syntactic structure: Sluicing and ellipsis in real-time sentence processing. *Language and Cognitive Processes, 28*(3), 272–302. https://doi.org/10.1080/01690965.2011.622905

Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*. https://doi.org/https://doi.org/10.17605/OSF.IO/MD832