CrossMark

The division of labor in explanations of verb phrase ellipsis

Christina S. Kim[1] · Jeffrey T. Runner[2]

Abstract In this paper, we will argue that, of the various grammatical and discourse constraints that affect acceptability in verb phrase ellipsis (VPE), only the structural parallelism constraint is unique to VPE. We outline (previously noted) systematic problems that arise for classical structural accounts of VPE resolution, and discuss efforts in recent research on VPE to reduce explanations of acceptability in VPE to general well-formedness constraints at the level of information structure (e.g. Kehler in Linguist Philos 23(6):533–575, 2000; Coherence, reference and the theory of grammar, CSLI Publications, Stanford, 2002; Proceedings of semantics and linguistic theory, vol 25, 2015; Kertz in Language 89(3):390–428, 2013). In two magnitude estimation experiments, we show that—in line with Kehler's predictions—degradation due to structural mismatch is modulated by coherence relation. On the other hand, we consistently find residual structural mismatch effects, suggesting that the interpretation of VPE is sensitive to structural features of the VPE antecedent. We propose that a structural constraint licenses VPE, but that sentences violating this constraint can nevertheless be interpreted. The variability in acceptability is accounted for not by additional constraints on VPE in the grammar, but by the numerous general biases that affect sentence and discourse well-formedness, such as information structural constraints (as proposed by Kertz 2013), discourse coherence relations (Kehler 2000), sensitivity to Question Under Discussion structure (e.g. Ginzburg and Sag in English

✉ Christina S. Kim
c.s.kim@kent.ac.uk

Jeffrey T. Runner
Jeffrey.Runner@rochester.edu

[1] Department of English Language and Linguistics, School of European Culture and Languages, University of Kent, Cornwallis North West, Canterbury, Kent CT2 7NF, UK

[2] Department of Linguistics, University of Rochester, 503 Lattimore Hall, Rochester, NY 14627, USA

🖄 Springer

interrogative constructions, CSLI Publications, Stanford, 2000; Kehler 2015), and thematic role bias at the lexical level (e.g. McRae et al. in J Mem Lang 38:283–312, 1998). We test the prediction that thematic role bias (Experiment 3) and QUD structure (Experiment 4) will influence both elliptical and non-elliptical sentences alike, while structural mismatch continues to degrade elliptical sentences alone. Our proposal differs from existing proposals in cutting the explanatory pie in a different way with respect to how variations in acceptability are accounted for. We suggest that degradation can result from at least two distinct and separable sources: violating construction-specific grammatical constraints, or from complexity differences in interpretation related to very general discourse level information.

1 Introduction

Sentences containing verb phrase ellipsis are characterized by a missing VP, as in the second clause in (1a). Despite the absence of any overt VP material, (1a) can only be interpreted as (1b), not as (for example) (1c).

(1)   a.   Christina emailed Mike, and Jeff did _____ too.
      b.   Christina emailed Mike, and Jeff emailed Mike.
      c.   Christina emailed Mike, and Jeff called Christina.

The classical explanation is that the grammar places restrictions on the environments in which VPE can occur.[1] Specifically, the elided VP must have a local antecedent VP ['emailed Mike' in (1a)], and these VPs are subject to some kind of identity constraint (Sag 1976; Williams 1977; Sag and Hankamer 1984; Dalrymple et al. 1991; Fiengo and May 1994). This rules out (1c) as a possible interpretation for (1a), since 'called Christina' does not occur anywhere in the preceding clause.

Consistent with an identity restriction on VPE, (2b), where the antecedent and ellipsis clauses are mismatched in voice, seems ill-formed, or at least degraded, relative to (2a), where the antecedent and ellipsis clauses are structurally parallel (both are in active voice).[2]

(2)   a.   Christina emailed Mike, and Jeff did, too.
      b.   ?Mike was emailed by Christina, and Jeff did, too.

While the existence of some type of identity constraint on VPE is generally agreed upon, the exact nature of this restriction has long been under debate. Much of the

---

[1] VPE is sometimes characterized as *post-auxiliary ellipsis*, due to examples like *Jeff is taller than Mike, and Chris is _____ too*, where the elided element may not be a VP (Miller and Pullum 2013).

[2] Throughout, we use *identity* and *syntactic/structural parallelism* interchangeably; e.g. (2a) satisfies the identity condition by virtue of the two clauses being structurally parallel, whereas (2b) violates identity because the antecedent and ellipsis clauses are not structurally parallel. We return to the question of exactly how to characterize the parallelism requirement below in Sects. 3 and 5.2.

disagreement has centered around two questions. The first is about the *generality* of the explanation: is the restriction specific to VPE, or does it apply to all sentences/discourses, with VPE simply showing one instance of a much broader phenomenon? The second has to do with *level of representation*: At what level of linguistic representation (syntax, semantics, information structure, among other possibilities) must the identity constraint hold? While proponents of a structural identity constraint (Hankamer and Sag 1976; Williams 1977; Fiengo and May 1994) have argued for a structural identity condition on the basis of degradation resulting from structural non-identity, as in (2b), others have argued that elided VPs are proforms, lacking structural content and taking any semantically-matched element as an antecedent (Dalrymple et al. 1991; Hardt 1993, 1999; Shieber et al. 1996; Hardt and Romero 2004).

Part of why a unified account of VPE has been elusive is that, as many researchers have noted, the pattern of acceptability judgments is strikingly graded. If the identity constraint is assumed to be categorical (either satisfied or violated, with nothing in between), sentences like (2b) are predicted to be simply ungrammatical. However, not all instances of identity-violating VPE appear to be equal. As noted by Hardt (1993), Kehler (2000), Kennedy and Merchant (2000) and Arregui et al. (2006), among others, the relative weakness or absence of structural mismatch effects in (3) suggests that, at least under certain conditions, strict structural identity may not be required.[3]

(3) a. This information could have been released by Gorbachov, but he chose not to ~~release it~~. *(Daniel Shorr, NPR, 10/17/92, from Hardt 1993)*
     b. In March, four fireworks manufacturers asked that the decision be reversed, and on Monday the ICC did ~~reverse it~~. *(from Rosenthal 1988; cited in Dalrymple 1991, Kehler 2002)*
     c. This problem was to have been looked into, but nobody did ~~look into it~~. *(from Kehler 2002)*

In this paper, we present empirical findings that bear on both the generality question and the level of representation question. We show that the acceptability of sentences with VPE is affected by mismatches or clashes at various levels of representation (lexical, syntactic, and discourse), each of which has previously been argued to be the locus of (non-)identity effects for VPE. However, only the syntactic structural mismatch effects are shown to be exclusive to VPE. We construe this as evidence for a VPE-specific structural identity condition, along the lines proposed by e.g. Hankamer and Sag (1976), though we remain open to alternatives that may superficially resemble structural identity. By delegating different sources of (un)acceptability to their appropriate sources—some general, and some construction-specific—this data sheds light on the generality question, and why there has been so much disagreement among researchers about the basic acceptability data itself.

---

[3] We indicate elided material in example sentences using strikethrough; where examples taken from other sources, differences in notation were modified for the sake of uniformity.

1.1 Attempts to derive the identity constraint from general principles

One family of theories has attempted to provide an explanation for the pattern of acceptability in VPE based on general discourse-level principles (Kehler 2002; Kertz 2013; also see Ginzburg and Sag 2000; Jacobson 2014; Kehler 2015, for related accounts). According to such accounts, what appears to be a construction-specific constraint is actually a reflex of a general constraint concerning coherence relations between sentences (Kehler 2002), or concerning information structural requirements (Kertz 2013). Such a theory would be quite attractive for a number of reasons. For one, it would be more parsimonious: rather than positing a distinct set of constraints for each of a number of ellipsis constructions, the patterns of acceptability would be accounted for by independently motivated aspects of the grammar. In addition, such a theory would be more explanatory: rather than a seemingly arbitrary restriction on the environments in which VPE is licensed, information structure or coherence structure would be able to systematically predict the discourse environments where VPE should be possible.

For Kehler (2000, 2002), VPE resolution is a product of the process of recognizing and establishing coherence relations in the course of discourse processing. He explains the variable sensitivity to structural match by allowing both semantic and syntactic mechanisms for recovering the elided VP under different circumstances. Which ellipsis resolution mechanism is used is determined by the coherence relation relating the antecedent and ellipsis clause, and because these mechanisms differ in their sensitivity to strict structural identity, whether structural mismatch counts as violating the relevant identity condition depends on the coherence relation. Specifically, an elided expression in a sentence that is part of a CAUSE-EFFECT relation only needs to match its antecedent in propositional content, and is not expected to require structural information. On the other hand, the elided material in a RESEMBLANCE relation relies on aligning its syntactic arguments with those of its antecedent, and therefore should show degradation when there is structural mismatch.

Coherence-based accounts predict interactions between Mismatch and Discourse relation type. The sentences in (4) should be worse than their Matched counterparts in (5), since they are instances of conjuncts in a RESEMBLANCE relation.[4]

(4)   a. *Active antecedent + Passive ellipsis* [RESEMBLANCE]:
        John implemented the computer system with a manager, but it wasn't ~~implemented with a manager~~ by Fred. (*Example (34) from Kehler 2000*)

      b. *Passive antecedent + Active ellipsis* [RESEMBLANCE]:
        This problem was looked into by John, and Bob did ~~look into the problem~~ too. (*Example (34) from Kehler 2000*)

(5)   a. *Active antecedent + Active ellipsis* [RESEMBLANCE]:
        John implemented the computer system with a manager, but Fred didn't ~~implement it with a manager~~.

---

[4] These are examples of Kehler (2000)'s PARALLEL relation, which he specifies as a subtype of RESEMBLANCE. To avoid confusion with other uses of *parallel* we refer to the discourse relation as RESEMBLANCE throughout.

    b.   *Passive antecedent + Passive ellipsis* [RESEMBLANCE]:
        The first problem was looked into by John, and the second problem was too
        ~~looked into by John~~.

However (6) should not be worse than (7) because they are instances of the CAUSE-EFFECT relation. Ellipsis resolution in such cases should involve a mechanism like higher-order unification (Dalrymple et al. 1991), which does not access the internal syntactic structure of the clause.

(6)   a.   *Active antecedent + Passive ellipsis* [CAUSE-EFFECT]:
        Actually I have implemented it [= a computer system] with a manager,
        but it doesn't have to be ~~implemented with a manager~~. (*Example (24) from*
        Kehler 2000)
    b.   *Passive antecedent + Active ellipsis* [CAUSE-EFFECT]:
        This problem was to have been looked into, but obviously nobody did ~~look~~
        ~~into the problem~~. (*Example (22) from Kehler 2000*)

(7)   a.   *Active antecedent + Active ellipsis* [CAUSE-EFFECT]:
        Actually I have implemented it with a manager, but I didn't have to
        ~~implement it with a manager~~.
    b.   *Passive antecedent + Passive ellipsis* [CAUSE-EFFECT]:
        This problem was to have been looked into, but obviously it wasn't ~~looked~~
        ~~into~~.

Kehler's theory makes concrete predictions about which cases of structural mismatch should be acceptable. The strongest form of his proposal predicts sensitivity to structural mismatch under the RESEMBLANCE relation, and complete insensitivity to syntactic structure under CAUSE-EFFECT.[5] In addition, because the theory is not specific to VPE, it predicts that non-elliptical sentences should show an analogous sensitivity to structural mismatch contingent on coherence relation.

Another strategy for addressing the pattern of acceptability in VPE has been to recharacterize the classical structural identity constraint as falling out from general well-formedness conditions on information structure representations (Hendriks 2004; Kertz 2013; Miller and Pullum 2013).[6] According to this family of theories, VPE and other varieties of ellipsis are consistent with particular information structure representations; if those representational expectations are not met, ellipsis is degraded or not possible.

When these various cues to information structure are incongruent—that is, when there is conflict between the information structural properties of VPE and other focusing devices—acceptability is degraded. Unlike the classical identity-based theories, these accounts say nothing specifically about ellipsis: the information structural prop-

---

[5] It is not clear whether Kehler intends this strictest view, but we spell it out here and below in Sect. 2.4 in order to generate explicit, testable predicitions.

[6] Ginzburg and Sag (2000) and Ginzburg (2012) give general accounts of ellipsis that rely on the notion of Question Under Discussion, focusing on fragment answers and sluicing. We will return to relevance of QUD for VPE below in Sect. 4.

erties of VPE are determined by the focusing effect of eliding some elements but not others, not by any VPE-specific condition. Rather, when various linguistic cues relevant for interpretation conflict, there is greater uncertainty about the final interpretation, resulting in decreased acceptability. Unacceptability related to mismatch in VPE is simply a specific case that results in an information structure clash at the discourse level. Such theories are particularly attractive because the theory independently needs information structural constraints to explain what constitutes discourse-level well-formedness, and no additional construction-specific constraint or licensing condition would be required.

These accounts differ from each other in a number of respects, but have in common that they place an alignment or identity constraint on the topic-focus structure of the antecedent and ellipsis clauses. Both Hendriks (2004) and Kertz (2013) present alternatives to Kehler's coherence account of ellipsis. Kertz (2013) proposes a single information structural alignment constraint on VPE, given below in (8).

(8) *Constraint on Contrastive Topic Relations* (Kertz 2013):
    A contrastive topic relation is well formed if members of the topic set are sentence topics.

Like semantic identity accounts, and Kehler's account, Kertz's theory predicts that structural mismatch should be acceptable under the right conditions. In (9) the contrastive topic 'poisonous plants' is contrasted with 'venomous snakes.' Because 'venomous snakes' is included in the topic set (for (9), {'venomous snakes', 'poisonous plants'}) by virtue of being interpreted as contrastive with 'poisonous plants', but is not a sentence topic, the constraint in (8) is violated, and this is predicted to degrade acceptability.

(9) It's easy to identify venomous snakes, and poisonous plants are as well.
(10) Venomous snakes are easy to identify, and most experienced hikers can.

By contrast, (10) has only simple focus: the two arguments ('venomous snakes' and 'most experienced hikers') are not in a contrastive topic relation, so the Constraint on Contrastive Topic Relations is satisfied and the mismatch is predicted not to affect acceptability.

However, contrast alignment does not seem to be the whole story: while mismatch in simple focus VPE is judged more acceptable than mismatch in contrastive topic VPE, Kertz's own data show that structural mismatch systematically degrades acceptability, regardless of focus structure: (9) and (10) are degraded relative to their structurally matching counterparts, (11) and (12).

(11) Venomous snakes are easy to identify, and poisonous plants are as well.
(12) It's easy to identify venomous snakes, and most experienced hikers can.

Thus, while showing that violating the focus structural constraint influences VPE well-formedness when syntactic structure is held constant, Kertz has also shown that structural mismatch influences well-formedness when the focus constraint is satisfied across the board.

Hendriks (2004) argues that Kehler's RESEMBLANCE relation actually relies on contrastive topics, with the two clauses linked to an implicit question. While Hendriks does not explicitly offer an analysis of mismatch in VPE, it would seem that degradation arises because—unlike in matched VPE, with contrastive topics—there is no coherent implicit question that can be linked to both clauses. Because clauses related by CAUSE-EFFECT are linked to two independent questions in Hendriks' account, any across-the-board degradation resulting from structural mismatch would be unexpected.

As noted above, a discourse or information-structural account of VPE would be more elegant than one that posits a seemingly arbitrary constraint on a restricted class of sentences. The question is whether such an account is empirically supported.

1.2 What a theory of VPE should and shouldn't account for

Because general accounts are not strictly about ellipsis (including VPE), but rather attempt to derive apparent constraints on VPE from discourse well-formedness, they make the same predictions about mismatch-related degradation in non-elliptical sentences as in sentences containing VPE. (13b) seems less acceptable than (13a), even though neither sentence contains VPE (examples from Kertz 2013, Table 4/Experiment 2).

(13) a. Venomous snakes are easy to identify, and most experienced hikers can identify them.
      b. ?It's easy to identify venomous snakes, and poisonous plants are easy to identify as well.

We agree that auxiliary-focus [as in (10) and (13a)], certain coherence relations [as in (6)], and other discourse-level constraints (e.g. Ginzburg and Sag 2000; Frazier and Clifton 2006; Grant et al. 2012; Jacobson 2014) improve acceptability in sentences with VPE. However, the experiments we present here show that even manipulating the discourse to be favorable to VPE does not eliminate a small but detectable degradation related to structural mismatch. We claim that the structural identity constraint responsible for this is the only identity condition that applies specifically to VPE. The small structural mismatch penalty observed in the more acceptable versions of VPE (auxiliary-focus, CAUSE-EFFECT relation) therefore represents the 'best case' scenario for mismatched VPE, where structural identity is violated but all other general constraints are satisfied. By contrast, the larger penalty observed in sentences with contrastive subjects or RESEMBLANCE is the cumulative penalty of non-VPE-specific factors and the VPE-specific structural constraint. More generally, the cumulative effects of construction-specific and general factors on acceptability create a very graded-looking distribution of acceptability judgments. In addition, we find the same acceptability pattern whether the antecedent and ellipsis clauses are coordinated or in separate sentences, leading us to suggest that this structural constraint operates at a level broader than sub-sentential syntax.

Because we see the goal of ellipsis resolution to be recovering the meaning of the elided constituent [cf. the Recoverability Condition on ellipsis (Katz and Postal

1964)], a sentence containing VPE can receive an interpretation despite violating exact structural identity. However, a less than perfect match can make VPE *harder* to interpret (i.e. recover a meaning for), and therefore degraded in terms of acceptability (for other proposals linking processing complexity with degraded acceptability, see Arregui et al. 2006; Hofmeister et al. 2013). We refer to this as a structural constraint, since it essentially builds into the semantic representation information about voice alternations, which we take to be a canonical syntactic alternation. But insofar as there is evidence for treating e.g. active and passive semantic representations differently, this constraint could be recast as an entirely semantic one.

We begin in Sect. 2 by testing the predictions of Kehler (2002), asking how mismatch effects are modulated by discourse coherence—in particular, whether the coherence effects reflect a VPE-specific mechanism, or general biases in interpretation. In Sect. 3, we ask to what extent mismatch effects are localized to contexts in which the antecedent appears in a conjoined clause, or also characterize VPE in contexts where the antecedent appears in a separate sentence. Section 4 begins to explore lexical and discourse-level pressures that likely conspire to cause a comprehender to arrive at a particular interpretation of a clause containing VPE. Based on the persistence of the structural mismatch effect, we suggest that a general, discourse or information structural theory cannot fully account for the pattern of acceptability in VPE. We conclude by considering some ways of making existing formal accounts compatible with our empirical findings.

2 Experiment 1: Structural effects are modulated by discourse coherence

Experiment 1 compares VPE sentences with matched or mismatched voice, where the discourse relation between the antecedent and the ellipsis conjuncts was either RESEMBLANCE or CAUSE-EFFECT. According to Kehler (2000, 2002), the RESEMBLANCE relation relies on the alignment of arguments from one sentence (or conjunct) to the next, and as such should be sensitive to changes in structural parallelism. On the other hand, the CAUSE-EFFECT relation relates sentences at the propositional meaning level, and should therefore be insensitive to structural manipulations that leave the meaning contribution of each conjunct intact. If discourse coherence is responsible for determining sensitivity to structural mismatch, the same acceptability pattern is expected for VPE sentences and their non-elliptical counterparts, with only RESEMBLANCE showing degradation when structural identity is violated.

Frazier and Clifton (2006) were the first to test Kehler (2000, 2002)'s discourse coherence approach. They used active VPE sentences with passive antecedents and connectives like 'and' or 'just like' to indicate a RESEMBLANCE relation, and 'because' to indicate a CAUSE-EFFECT relation. In both percent 'got it' data and acceptability ratings on a 1-to-5 scale, they failed to find the asymmetry predicted by Kehler: the RESEMBLANCE cases were no worse when clauses were mismatched in voice than the CAUSE-EFFECT cases. While they did show for the acceptability ratings that corresponding sentences with structurally matching conjuncts did not differ—that is, the lack of difference in the mismatching sentences was not due to a difference in the acceptability of the matching forms—they did not show this for unelided versions of

their items. This is potentially problematic, since it is impossible to determine whether unelided sentences have different baseline acceptabilities. For instance, it is plausible that the causal connective 'because' carries additional temporal implications that may be violated by the clauses being in the same tense, in a way that analogous items with connectives 'and' or 'just like' are not susceptible to. Such an asymmetry would skew results in the direction of mismatched CAUSE-EFFECT conditions being worse, reducing the chances of finding a RESEMBLANCE/CAUSE-EFFECT contrast. Indeed, Kertz (2008) tested the possibility of base-line differences by embedding Frazier and Clifton (2006)'s original materials in a magnitude estimation experiment that included unelided controls. Kertz (2008) found that Frazier and Clifton (2006)'s finding—that CAUSE-EFFECT sentences were worse overall than RESEMBLANCE sentences—was true of both the VPE sentences and their unelided counterparts. This contrasts with another magnitude estimation study by Kim et al. (2011), which tested both voice mismatch and category mismatch (nominal antecedents) in VPE and unelided controls, and found a mismatch penalty only in sentences containing ellipsis.

To provide a more complete test of Kehler's predictions about structural parallelism, the current experiment includes both structural match and no-ellipsis conditions, along with the two coherence conditions. The match conditions, where no degradation is predicted, provide appropriate baselines for the corresponding mismatch conditions, allowing us to observe whether and to what extent the coherence relation modulates mismatch-related degradation. Comparison of VPE with no-ellipsis conditions will show whether any coherence effects are limited to sentences requiring VPE resolution.

2.1 Design

There were 8 conditions in the experiment [Ellipsis (Ellipsis, No Ellipsis) × Mismatch (Match, Mismatch) × Discourse Relation (RESEMBLANCE, CAUSE-EFFECT)]. Half of the Match trials had active VPs, and half had passive VPs. Similarly, half of the Mismatch trials had Active–Passive order, and the other half had Passive–Active order.

We constructed two separate sets of items for Resemblance and Cause-Effect, in order to optimize the acceptability of the Match and No Ellipsis conditions. Since we are primarily interested in changes in acceptability associated with mismatch, and whether this is conditioned on the presence of ellipsis, we considered it important to have the baseline (i.e. No Ellipsis, Match) conditions sound as natural as possible. This can be quite difficult when the same set of items is used for different coherence relations (see e.g. Frazier and Clifton 2006). Note also that Kehler (2002) cautions that simply changing a connective might not be sufficient to change the coherence relation signalled: in addition to the connective, the relation is cued by the propositional content of the two clauses, and the likelihood of e.g. a causal relation holding between the two.

As illustrated in the example stimuli in (14)–(15), the connectives 'so' or 'because' were used for CAUSE-EFFECT items, and the connectives 'and' or 'but' as well as the modifier 'too' after the ellipsis site were used for RESEMBLANCE.

(14)   RESEMBLANCE *conditions*

    a. *No Ellipsis, Match:*
        The crime lab analyzed the blood sample, and the private clinic analyzed it, too.

    b. *No Ellipsis, Mismatch:*
        The crime lab analyzed the blood sample, and the fingerprints were analyzed, too.

    c. *Ellipsis, Match:*
        The crime lab analyzed the blood sample, and the private clinic did, too.

    d. *Ellipsis, Mismatch:*
        The crime lab analyzed the blood sample, and the fingerprints were, too.

(15)   CAUSE-EFFECT *conditions*

    a. *No Ellipsis, Match:*
        Sue had requested that the presenters dim the lights during the announcement, so they turned the lights down.

    b. *No Ellipsis, Mismatch:*
        Sue had requested that the presenters dim the lights during the announcement, so they were turned down.

    c. *Ellipsis, Match:*
        Sue had requested that the presenters dim the lights during the announcement, so they did.

    d. *Ellipsis, Mismatch:*
        Sue had requested that the presenters dim the lights during the announcement, so they were.

2.2 Method and procedure

The experimental paradigm used was magnitude estimation, adapted from e.g. Bard et al. (1996). In this paradigm, participants give numerical ratings to stimuli relative to the rating they gave to some standard, or modulus, at the beginning of the experiment.[7] For language stimuli, the ratings are participants' estimates of the acceptability of the sentence in the current trial compared to the acceptability of the modulus.

The current experiment was run on a Macintosh computer running PsyScope X software (Bonatti 2008). Participants first practiced giving estimates of line lengths (cf. Bard et al. 1996), then practiced with sentences. Then they assigned a value to the modulus sentence: *The children were amused by the cartoon, but their parents weren't.* On each trial, the modulus appeared on the screen, together with the sentence to be rated on that trial. Participants typed their estimates into a text box, then pressed the spacebar to proceed to the next trial. Experimental trials were interspersed with filler sentences, which were either monoclausal or contained a discourse relation that did

---

[7] Although Sprouse (2011) and Wescott and Fanselow (2011) critically examine magnitude estimation as a methodology for collecting acceptability judgments, Schütze and Sprouse (2014) ultimately conclude that it is no better or worse than other methods such as Likert and thermometer scales.

not appear in the test items (see "Appendix A"; the instructions used for Experiments 1–2 are given in "Appendix B").

There was one break halfway through the trials, and the whole experimental session took participants approximately 15 min. 20 native English speakers from the University of Rochester community participated.
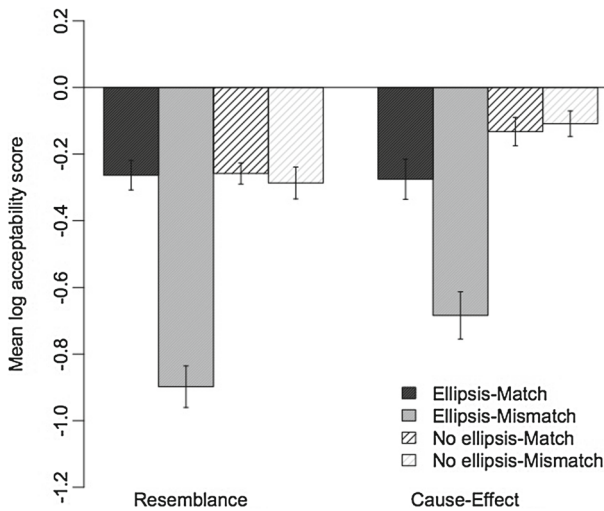
2.3 Results

2.3.1 Data analysis

The data were first normalized by dividing each participant's estimates by their modulus value. All analyses were performed on log-transformed values of the normalized data. Data for all experiments were trimmed at 4 standard deviations from the grand mean. All contrasts reported were significant at $p < 0.05$, and where appropriate in *post hoc* tests, corrections were made for multiple comparisons.

The log scores were fit to a linear mixed-effects model including Ellipsis, Mismatch, Discourse relation, and Antecedent voice as fixed effects (Baayen 2008). In all models reported, all factors manipulated in the experiment and all the interactions among these factors were included as predictors (i.e. predictors were not removed using model comparison). Throughout, we report linear mixed-effects models with the maximal random effects structure justified by the data (Barr et al. 2013). The procedure for determining the random effects structure was as follows. We started with the maximal random effects, then, if the model did not converge, removed random effect terms one at a time, starting with the highest order term, and starting with Item random effects within same-order terms. This procedure was iterated until the model converged. In each case, the final model is shown at the top of the table containing model coefficient estimates.

The condition means for Experiment 1 are plotted in Fig. 1. Note that the mean acceptability judgments for the experimental conditions were all below zero, indicating that test items were judged less acceptable overall than the modulus sentence. While we are primarily concerned with relative differences among test conditions, we verified that the acceptability ratings for the filler items did not make the test items stand out from the entire set of materials. The fillers ranged from -13.1 to 1.9 (log(Acceptability)); ungrammatical fillers were judged less acceptable than grammatical ones ($\beta = -1.02$, SE $= 0.12$, $p < 0.0001$), and there was a length-grammaticality interaction such that long fillers were less acceptable than short ones to a greater extent for ungrammatical fillers than grammatical ones ($\beta = 0.29$, SE $= 0.13$, $p < 0.05$).

2.3.2 Main effects and interactions

There were main effects of Ellipsis, Mismatch, and Discourse relation. Sentences containing ellipsis were judged less acceptable than their counterparts without ellipsis, and sentences with mismatching clauses were judged worse than their matching counterparts. In addition, sentences where the two clauses were related by the RESEMBLANCE relation were judged less acceptable than those with CAUSE-EFFECT.

Fig. 1 Experiment 1 results (*error bars* represent standard error)

There was also an Ellipsis-Mismatch interaction, such that mismatched sentences were judged less acceptable than matching counterparts in sentences with ellipsis, but not in unelided sentences. In addition, Mismatch interacted with Discourse relation—the Mismatch effect was greater in RESEMBLANCE conditions than in CAUSE-EFFECT conditions. There was a Mismatch-Antecedent voice interaction, such that in Mismatch conditions, Passive–Active order was judged more acceptable than Active–Passive order (we return to this point in Sect. 2.3.4). Finally, there was a three-way Ellipsis-Mismatch-Discourse relation interaction: the degradation associated with mismatch in ellipsis sentences was greater when the discourse relation was RESEMBLANCE than when it was CAUSE-EFFECT. Estimates of the model coefficients corresponding to the fixed effects are given in Table 1.[8]

While the mismatch penalty is decreased when the antecedent and ellipsis clauses are related by CAUSE-EFFECT (difference in means between Match-Ellipsis and Mismatch-Ellipsis conditions = 0.49) rather than RESEMBLANCE (difference in means = 0.76), it is not entirely eliminated: a planned comparison shows that even in CAUSE-EFFECT conditions, sentences with voice mismatch and ellipsis were degraded relative to matching sentences with ellipsis ($t = -5.17$, $p < 0.0001$).

---

[8] A quick primer on how to read tables with model coefficients. The leftmost column represents the predictors in the regression model; two-way and three-way interaction terms are represented as A:B and A:B:C. The β value for each term is the estimate of the coefficient for that term in the model. SE, t, and p are the standard error, t value and p value associated with the coefficient estimate. Thus, the Mismatch row indicates that there is a negative effect of Mismatch, which is significant—in other words, when there is structural mismatch, the judgements of acceptability are significantly worse than when the two clauses match structurally. All predictors were contrast-coded. In all tables, significance level is indicated as follows: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, .$p < 1.0$, $n.s.p > 1.0$ (non-significant).

Table 1 Experiment 1 model and model coefficients

```
log(Acceptability) ~ Ellipsis + Mismatch + RESEMBLANCE + Antecedent
    + Ellipsis:Mismatch + Ellipsis:RESEMBLANCE + Ellipsis:Antecedent
    + Mismatch:RESEMBLANCE + Mismatch:Antecedent + RESEMBLANCE:Antecedent
    + Ellipsis:Mismatch:RESEMBLANCE + Ellipsis:Mismatch:Antecedent
    + Ellipsis:RESEMBLANCE:Antecedent + Mismatch:RESEMBLANCE:Antecedent
    + Ellipsis:Mismatch:RESEMBLANCE:Antecedent
    + (1 + Ellipsis + Mismatch + RESEMBLANCE + Antecedent|Subject)
    + (1 + Ellipsis + Mismatch + Antecedent|Item)
```

	β	SE	t	p
Intercept	−0.24	0.081	−2.99	**
Ellipsis	−0.14	0.027	−5.25	***
Mismatch	−0.13	0.025	−5.33	***
RESEMBLANCE	−0.11	0.025	−4.58	***
PassiveAntecedent	0.033	0.025	1.31	*n.s.*
Ellipsis:Mismatch	−0.13	0.017	−7.62	***
Ellipsis:RESEMBLANCE	−0.0087	0.017	−0.50	*n.s.*
Mismatch:RESEMBLANCE	−0.069	0.019	−3.72	***
Ellipsis:PassAntecedent	0.0017	0.018	0.096	*n.s.*
Mismatch:PassAntecedent	0.078	0.018	4.38	***
RESEMBLANCE:PassAntecedent	0.0073	0.017	0.43	*n.s.*
Ellipsis:Mismatch:RESEMBLANCE	−0.047	0.019	−2.48	*
Ellipsis:Mismatch:PassAnt	0.032	0.017	1.91	.
Ellipsis:RESEMBLANCE:PassAnt	0.030	0.019	1.59	*n.s.*
Mismatch:RESEMBLANCE:PassAnt	−0.015	0.018	−0.85	*n.s.*
Ellipsis:Mismatch:RESEMBLANCE:PassAnt	0.014	0.019	0.73	*n.s.*

2.3.3 Additional analyses of Experiment 1 data

The main point of Experiment 1 is to demonstrate the differential effect of coherence relation on ellipsis clauses involving matching or mismatching antecedents. A reviewer pointed out that, in a number of the items, within the CAUSE-EFFECT conditions, no ellipsis sentences featured non-identical clauses. One reason for non-identity was the lexical material in the second VP—including the verb itself—differing from the VP in the first clause, as in (16).

(16)   a.   The class requested that the exam be rescheduled, so it was./was moved to a different time.

          b.   The class requested that the exam be rescheduled by the professor, so she did./moved it to another time.

An additional subset of CAUSE-EFFECT-No ellipsis items involved additional material in the second clause as in (17)—typically a prepositional phrase or a temporal modifier.

(17) a. Andrea asked Jim to turn on the heat in their apartment, so he did./turned it on when he got home.

   b. Andrea asked Jim to turn on the heat in their apartment, so it was./was turned on when he got home.

This subset of items might be problematic in terms of interpreting the data, since the unelided sentences involve different propositions than their elided counterparts, for which the only antecedent possible is the VP in the antecedent clause (or a voice-mismatched version of that VP).[9]

In order to determine whether the above subset of items was responsible for the results of Experiment 1, we tagged the items as having any of the following character-istics: (1) different verbs used in antecedent/ellipsis clauses, (2) modifiers (e.g. PPs, temporal modifiers) following ellipsis, or (3) long passive as opposed to short passive (see e.g. Mauner et al. 1995, for suggestions that short and long passives may differ in their sensitivity to parallelism, at least for deep anaphors), and excluded these from the dataset. A model with the same predictors as the model in the main text shows the same effects when items having any of the above characteristics were excluded (Table 2).

In addition, our primary objective in Experiment 1 was to demonstrate that within Ellipsis conditions, there was a greater effect of Mismatch for RESEMBLANCE than there was for CAUSE-EFFECT. (Note that the issues described above generally do not apply to the Ellipsis conditions, as the concerns were about differences between Ellipsis and No Ellipsis conditions.) To confirm that this result would still be obtained, we fitted just the Ellipsis portion of the dataset with a model including all the same predictors as the above model except for Ellipsis, again excluding any potentially problematic items. This model (Table 3) confirms that the problematic properties pointed out by the reviewer were not responsible for carrying the results we report.

2.3.4 Comparison with Arregui et al. (2006)

Our findings confirm one interesting aspect of Arregui et al. (2006): we also find that Passive–Active order is judged more acceptable than Active–Passive order, across sentences with and without ellipsis. This is reflected in our model as an interaction between Antecedent voice and Mismatch. Arregui et al. (2006) explain their effect in terms of markedness: because passive voice is the marked form, it is more susceptible to be misremembered as the unmarked active form in mismatched VPE (Passive–

---

[9] An additional item (i) (item 3, "Appendix A") was potentially problematic, because the verbs in the active (ia) and passive (ib) do not have identical lexical meanings.

(i) a. Bruce burned himself on the stove, and Jason did, too./burned himself, too.
   b. Bruce burned himself on the stove, but Jason wasn't./wasn't burned.

The sentence in (ia), with a bound reflexive, is normally interpreted as without an agent argument (e.g. *Bruce burned himself accidentally*). By contrast, (ib) is most easily interpreted as having an implicit agent; it also seems incompatible with an explicit 'by himself' *by*-phrase (*Bruce was burned by himself). This item was also excluded from the analysis in Table 2.

Table 2 Experiment 1 data subset model and model coefficients

```
log(Acceptability) ~ Ellipsis + Mismatch + RESEMBLANCE + Antecedent
    + Ellipsis:Mismatch + Ellipsis:RESEMBLANCE + Ellipsis:Antecedent
    + Mismatch:RESEMBLANCE + Mismatch:Antecedent + RESEMBLANCE:Antecedent
    + Ellipsis:Mismatch:RESEMBLANCE + Ellipsis:Mismatch:Antecedent
    + Ellipsis:RESEMBLANCE:Antecedent + Mismatch:RESEMBLANCE:Antecedent
    + Ellipsis:Mismatch:RESEMBLANCE:Antecedent
    + (1 + Ellipsis + Mismatch + RESEMBLANCE|Subject)
    + (1 + Ellipsis + Mismatch + Antecedent|Item)
```

	β	SE	t	p
Intercept	−0.24	0.079	−3.07	**
Ellipsis	−0.14	0.031	−4.60	***
Mismatch	−0.14	0.031	−4.45	***
RESEMBLANCE	−0.11	0.027	−4.15	***
PassiveAntecedent	0.030	0.023	1.31	*n.s.*
Ellipsis:Mismatch	−0.14	0.018	−7.73	***
Ellipsis:RESEMBLANCE	0.0097	0.018	−0.54	*n.s.*
Mismatch:RESEMBLANCE	−0.072	0.019	−3.73	***
Ellipsis:PassAntecedent	−0.0019	0.018	−0.11	*n.s.*
Mismatch:PassAntecedent	0.078	0.018	4.26	***
RESEMBLANCE:PassAntecedent	0.0031	0.017	0.18	*n.s.*
Ellipsis:Mismatch:RESEMBLANCE	−0.052	0.020	−2.64	**
Ellipsis:Mismatch:PassAnt	0.035	0.017	2.05	*
Ellipsis:RESEMBLANCE:PassAnt	0.025	0.019	1.31	*n.s.*
Mismatch:RESEMBLANCE:PassAnt	−0.014	0.018	−0.76	*n.s.*
Ellipsis:Mismatch:RESEMBLANCE:PassAnt	0.014	0.019	0.70	*n.s.*

Active order) than active voice is to be misremembered as passive (Active–Passive order).

2.3.5 Comparison with Frazier and Clifton (2006)

Another question brought up by these results is why Frazier and Clifton (2006) failed to find any coherence-related effects on acceptability, in their similar study manipulating voice mismatch in VPE, while our data show an Ellipsis-Mismatch-Discourse relation interaction. Recall that all of the structural mismatch stimuli in Experiments 1 and 2 of Frazier and Clifton (2006) have Passive–Active order. The Ellipsis-Mismatch-Discourse relation-Antecedent voice predictor is not significant in our regression model. However, when we analyzed only the Passive–Active sentences in our data, the Ellipsis-Mismatch-Discourse relation interaction was no longer a significant predictor ($\beta = -0.019$, $SE = 0.026$, $p > 0.1$). By contrast, the interaction remained significant in the Active–Passive subset of our data ($\beta = -0.063$, $SE = 0.029$, $p < 0.05$).

Table 3 Experiment 1 model and model coefficients (ellipsis conditions only)

```
log(Acceptability) ~ Mismatch + RESEMBLANCE + Antecedent
    + Mismatch:RESEMBLANCE + Mismatch:Antecedent
    + RESEMBLANCE:Antecedent + Mismatch:RESEMBLANCE:Antecedent
    + (1 + Ellipsis + Mismatch + RESEMBLANCE|Subject)
    + (1 + Ellipsis + Mismatch + Antecedent|Item)
```

	β	SE	t	p
Intercept	−0.36	0.091	−3.93	**
Mismatch	−0.28	0.054	−5.20	***
RESEMBLANCE	−0.12	0.034	−3.35	**
PassiveAntecedent	0.017	0.034	0.63	*n.s.*
Mismatch:RESEMBLANCE	−0.12	0.030	−3.84	***
Mismatch:PassAntecedent	0.086	0.028	3.03	**
RESEMBLANCE:PassAntecedent	0.030	0.028	1.10	*n.s.*
Mismatch:RESEMBLANCE:PassAnt	−0.010	0.029	−0.35	*n.s.*

It appears that, at least in our data, the modulation of the mismatch penalty on VPE by discourse coherence relied on mismatched sentences with Active–Passive order. Thus while we are still left with the interesting question of why reconstructing a passive from an active antecedent may be more sensitive to the discourse context in which violations of structural identity occur,[10] we may be able to reconcile the difference between the current findings and those reported in Frazier and Clifton (2006).[11]

2.4 Discussion

The results of Experiment 1 show that VPE is sensitive to structural parallelism as represented by the voice alternation, but that the extent of this sensitivity is modulated by the type of discourse coherence relation the antecedent and ellipsis conjuncts are part of—that is, the choice of coherence relation can make sentences with voice mismatched ellipsis sound more acceptable. However, a reliable mismatch penalty

---

[10] A reviewer pointed out that this might be related to the different levels of vP-complexity with actives and passive: according so some theories, active sentences contain an additional v-head which assigns ACC case. Our findings are consistent with an explanation along these lines, though such a theory does not necessarily uniquely explain our data.

[11] We note that Experiment 6 in Kertz (2010) also manipulates VPE and coherence within clauses, similarly to our Experiment 1. Whereas her study finds that, overall, CAUSE-EFFECT sentences containing ellipsis are degraded more than RESEMBLANCE ones, there is no interaction with Match (voice match/mismatch). In our Experiment 1 we do find such an interaction—RESEMBLANCE-Ellipsis-Mismatch is worse than CAUSE-EFFECT-Ellipsis-Mismatch—and that interaction is replicated in our Experiment 2. It may be a matter of power that Kertz's interaction does not reach significance. In addition, the mismatch effect may have been weakened because, like Frazier and Clifton (2006), Kertz's Experiment 6 uses exclusively Passive–Active order for voice mismatch. That we found the same results for both Experiments 1 and 2 suggests to us that the interaction is real.

remains even when the coherence relation provides a favorable environment for mismatched VPE. This resembles the pattern of results reported by Kim et al. (2011), with structural mismatch selectively degrading sentences with VPE. In addition, Experiment 1 shows that sentences featuring a CAUSE-EFFECT relation were less sensitive to structural mismatch in general—not only in sentences containing VPE.[12]

While these findings suggest that coherence alone cannot explain the entire pattern of acceptability in VPE, this relies on the RESEMBLANCE and CAUSE-EFFECT sentences from Experiment 1 reliably being construed as those coherence relations. We used connectives as the primary cues to a relation, but as Kehler (2000) has noted, just changing a connective might not be sufficient to change the relation being signalled. A specific potential concern relates to Hendriks (2004)'s proposal, according to which the presence of contrastive topics is what simulates the effects of a RESEMBLANCE relation. If we take the view that contrastive topics are responsible for the observed mismatch effects, it could be possible that any sentences in the CAUSE-EFFECT condition with contrastive topics were being interpreted more like those in the RESEMBLANCE condition. If this were the case, it could explain why mismatched CAUSE-EFFECT sentences were degraded relative to their matched counterparts. However, since the CAUSE-EFFECT sentences did not feature contrastive subjects (see "Appendix A"; note that many of the CAUSE-EFFECT items have pronominal subjects in the second clause, indicating non-contrastiveness with the subject of the first clause), it seems that both Hendriks and Kehler would predict acceptability to be unaffected by voice mismatch in this condition—Hendriks because each clause should have its own implicit question, and Kehler because the CAUSE-EFFECT relation joins clauses at the propositional level.

We return to the question of whether our CAUSE-EFFECT items actually convey true CAUSE-EFFECT relations in Experiment 4, where we directly manipulate the question that VPE sentences are meant to answer. Here, we turn to another possible explanation for why Experiment 1 showed mismatch effects across both coherence relations.

By taking the results of Experiment 1 to support Kehler's proposal—that acceptability in VPE is modulated by coherence—we are making the assumption that coherence relations hold between clauses in a coordinate sentence just as they do between sentences. This seems to us an intuitive extension of Kehler's original proposal; indeed, other researchers (e.g. Hardt and Romero 2004) have taken a similar approach in assuming that discourse connectives link proposition-sized units both within and across

---

[12] While we characterize the results of Experiment 1 in terms of Kehler's theory of discourse coherence (i.e. what we intended to manipulate), the current data do not allow us to distinguish this from Frazier and Clifton (2006)'s alternative hypothesis, which invokes a notion of parallelism distinct from the parallelism introduced by a RESEMBLANCE coherence relation. Specifically, they suggest that the presupposition introduced by the sentence-final 'too' in many RESEMBLANCE sentences is a source of parallelism effects in VPE. If this is the case, it would still be an instance of non-structural aspects of a sentence modulating the severity of structural parallelism effects. We would therefore still be interested in whether such effects can entirely amnesty VPE from structural identity requirements. Since information carried by the connective and 'too' become available at different locations in the sentence, future experiments using online measures may help separate out the timecourses of these potentially different sources of parallelism.

sentence boundaries.[13] However, there may be reasons not to assume that coherence relations hold between clauses in the same way as they do between sentences.

A common assumption in syntax is that core syntactic operations and principles are confined to the sentence domain—for instance, when considering possibilities for linguistic coreference, candidate referents outside the current sentence are not in the appropriate structural relationship (e.g. c-command) with reflexive anaphors inside the sentence. Miltsakaki (2002) similarly claims that whether clauses are coordinated or in separate sentences affects pronoun resolution. Discourse structural considerations are instead often assumed to play a more important role in relating sentences to each other in terms of their informational organization. Under this view, while the internal syntactic structure of one sentence may not affect the interpretation of subsequent sentences, it may contribute to the discourse structure by establishing what is given, focused, topical or the question under discussion. This in turn influences subsequent interpretation.

While they do not dispute the possibility of sentence-internal discourse representations, Frazier and Clifton (2005) take the view that structural effects are limited in their domain to the sentence; once outside the sentence domain, only discourse-level interpretive constraints (based on extracting the main assertion from a sentence) apply. They compared the acceptability of sentences like (18a), where an elided VP and its antecedent are in a single sentence with coordination, with pairs of sentences like (18b), where the antecedent VP is in one sentence and an elided VP is in another.

(18) a. John said that Fred went to Europe and Mary did, too.
      b. John said that Fred went to Europe. Mary did, too.

Frazier and Clifton (2005) find that in the coordination condition, people are more likely to construe the 'go to Europe' VP as the antecedent of the elided VP, while in the two-sentence condition, they are less likely to consider it the antecedent.

If Frazier and Clifton are right that structural effects are confined to the sentence domain because comprehenders do not retain detailed syntactic representations across sentences, the structural mismatch effects observed in Experiment 1 are predicted to go away if the dependency between an elided VP and its antecedent crosses a sentence boundary. However, following the basic intuition in Kehler's work—that the relations between meanings in a discourse context influence whether structural or semantic information is important for interpretation—we might expect such discourse-modulation to hold as much, if not more, across discourse units. Experiment 2 is designed to test these diverging predictions, and in doing so, to more precisely characterize the level of structural representation at which the mismatch effect is located.

3 Experiment 2: Structural effects above the 'syntactic' domain

Experiment 2 compares cases of VPE in a coordinate structure (Within-sentence) with VPE across a sentence boundary (Cross-sentential). Antecedent and ellipsis clauses

---

[13] See Prasad et al. (2008), which assumes that discourse connectives link proposition-sized units both within and across sentences; also Rohde et al. (2011), who report coherence-related effects across clauses not linked by an explicit connective.

were related either by a RESEMBLANCE or a CAUSE-EFFECT relation, as in Experiment 1. The Within-sentence conditions are identical in structure to the items in Experiment 1; as such, we expect to see the same Ellipsis-Mismatch interaction pattern in the current experiment. The critical question is whether this acceptability pattern is replicated in cross-sentential VPE, with VPE-specific degradation due to voice mismatch and a non-VPE-specific coherence effect.

3.1 Design

There were 8 conditions in the experiment [Ellipsis type (Within-sentence, Cross-sentential) × Discourse Relation (RESEMBLANCE, CAUSE-EFFECT) × Mismatch (Match, Mismatch)]. All of the Mismatch trials had voice mismatch; half of these had Active–Passive order, and half had Passive–Active order. All sentences contained ellipsis.[14] The design of Experiment 2 and example stimuli are given in (19)–(20).

(19) *Match conditions*

    a. *Coordination,* RESEMBLANCE:
        The crime lab analyzed the blood sample, and the private clinic did, too.

    b. *Cross-sentential,* RESEMBLANCE:
        The crime lab analyzed the blood sample. The private clinic did, too.

    c. *Coordination,* CAUSE-EFFECT:
        Abby insisted that Bill get rid of the video tape, so he eventually did.

    d. *Cross-sentential,* CAUSE-EFFECT:
        Abby insisted that Bill get rid of the video tape. So he eventually did.

(20) *Mismatch conditions*

    a. *Coordination,* RESEMBLANCE:
        The crime lab analyzed the blood sample, and the fingerprints were, too.

    b. *Cross-sentential,* RESEMBLANCE:
        The crime lab analyzed the blood sample. The fingerprints were, too.

    c. *Coordination,* CAUSE-EFFECT:
        Abby insisted that Bill's video tape be destroyed, so he eventually did.

    d. *Cross-sentential,* CAUSE-EFFECT:
        Abby insisted that Bill's video tape be destroyed. So he eventually did.

3.2 Method and procedure

The procedure and method were the same as in Experiment 1. Approximately half of the practice and filler items contained two sentences. 14 native English speakers from the University of Rochester community participated. Each experimental session took approximately 15 min.

---

[14] For test and sample filler items, see Appendix A.2. in the preprint version of this paper at http://semanticsarchive.net/Archive/mFkMTc4Y/.
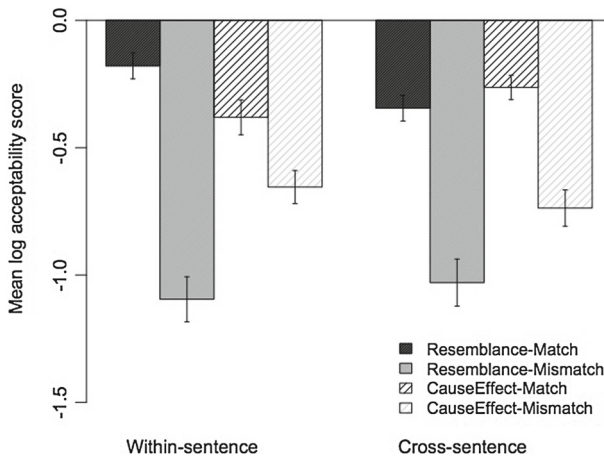
Fig. 2 Experiment 2 condition means (*error bars* are standard error)

3.3 Results

3.3.1 Data analysis

As before, data were normalized and log-transformed. Log scores were fit to a linear mixed-effects model with Ellipsis type, Discourse relation and Mismatch as fixed effects, and the maximal random effects structure justified by the data.

3.3.2 Main effects and interactions

As in Experiment 1, there was a main effect of Mismatch: mismatched sentences (or sentence pairs) were worse than their matching counterparts. There was also a marginal main effect of Discourse relation: RESEMBLANCE conditions were judged less acceptable than CAUSE-EFFECT conditions.

There was a Discourse Relation-Mismatch interaction, as in the previous experiment: sentences were more degraded under RESEMBLANCE than CAUSE-EFFECT, when there was mismatch between the two clauses, but not when the two clauses matched structurally. However, planned comparisons revealed that, as in Experiment 1, mismatched sentences were judged worse than matched ones even in CAUSE-EFFECT conditions ($t = 5.94$, $p < 0.0001$). No two-way or three-way interactions involving Ellipsis type were significant. In other words, whether VPE occurred within a single sentence or across two sentences had no effect on acceptability. Condition means are plotted in Fig. 2. Estimates of model coefficients are given in Table 4.

3.4 Discussion

Experiment 2 finds that the modulation of structural mismatch by discourse coherence when interpreting VPE (Experiment 1) extends to cases where the antecedent-ellipsis

Table 4 Experiment 2 model and model coefficients

```
log(Acceptability) ~ CrossSentence + Mismatch + RESEMBLANCE
      + CrossSentence:Mismatch + CrossSentence:RESEMBLANCE
      + Mismatch:RESEMBLANCE + CrossSentence:Mismatch:RESEMBLANCE
      + (1 + CrossSentence + Mismatch + RESEMBLANCE|Subject)
      + (1 + CrossSentence + Mismatch|Item)
```

	β	SE	t	p
Intercept	−0.59	0.091	−6.52	***
CrossSentence	−0.0053	0.038	−0.14	*n.s.*
Mismatch	−0.30	0.060	−5.01	***
RESEMBLANCE	−0.082	0.043	−1.88	.
CrossSentence:Mismatch	0.00456	0.034	0.16	*n.s.*
CrossSentence:RESEMBLANCE	−0.017	0.035	−0.48	*n.s.*
Mismatch:RESEMBLANCE	−0.11	0.035	−3.42	**
CrossSentence:Mismatch:RESEMBLANCE	0.053	0.035	1.51	*n.s.*

dependency is across sentences. This result is inconsistent with a categorical syntax-discourse divide like that proposed by Frazier and Clifton (2005): the Mismatch penalty was unaffected by whether antecedent and ellipsis site were in the same sentence, or separated by a sentence boundary.

The fact that we find that structural mismatch degrades VPE even across sentences suggests that language users have access to the kind of structural information relevant for VPE resolution at the level of discourse: as a sentence containing VPE is interpreted, the antecedent-ellipsis relation is constrained by structural parallelism irrespective of whether the antecedent is part of a single, connected syntactic structure. If sentence interpretation were limited in such a way that only one unit of syntactic structure could be attended to at a time, it would not be possible to compare the structure of an antecedent in a previous sentence to structure in the current sentence. In fact, it appears that any view of the relationship between syntactic and discourse structures where discourse representations contain no or very impoverished structural information will be unable to explain how structural identity can be enforced across discourse. The pattern of data from Experiment 2, where crossing a sentence boundary does not interact with either Coherence relation or voice mismatch, fits well with a view of discourse processing in which sentences constitute a meaningful unit of processing, but at least certain structural information can nonetheless persist across stretches of discourse that include more than one sentence unit. Additional evidence that mental representations of syntactic structure can persist after they have been processed come from structural priming studies (see e.g. Bock and Griffin 2000; Branigan et al. 2000; Gries 2005; Kaschak 2007) (though see Cai et al. 2012 for an opposing argument).

Our findings conflict with claims from the literature that, due to working memory limitations, such structural representations are only accessible for the current sentence (or some similarly restricted processing window; see e.g. Frazier and Clifton 2005). In

terms of level of representation, the structural parallelism constraint appears to operate at a level of representation that encodes broad structural differences (the difference between active and passive voice) both within and across sentence boundaries. While it is not itself an antecedent search mechanism, the structural identity constraint may interact with the process of locating candidate antecedents in particular ways. If an antecedent search mechanism (such as the content-addressable pointer mechanism proposed by Martin and McElree (2008), or the mechanism proposed by Arregui et al. (2006), which looks for candidate antecedents in canonical VP positions) identifies a number of candidate antecedents, the structural constraint may bias the comprehender toward one potential antecedent over another, based on the degree to which structural identity is satisfied or violated.

4 Experiments 3–4: Broad influences on discourse well-formedness

Experiments 1 and 2 have provided support for discourse coherence modulating effects of structural mismatch, as proposed by Kehler (2000). However, it appears that the discourse effects are broad effects, not specific to sentences containing ellipsis. In this respect, it may be similar to other lexical or discourse-level factors that may influence well-formedness in ellipsis and in non-elliptical sentences (for example, information structure, as proposed by Kertz 2013). On the other hand, Experiments 1 and 2 also show a persistent degradation due to structural mismatch, which *is* VPE-specific (see also Kim et al. 2011). Because the structural mismatch effect appears against a backdrop of multiple lexical and discourse-level biases which all contribute to a sentence's overall acceptability, looking at elliptical sentences alone may give the impression that such general pressures are additional constraints on ellipsis.

In the current section, we focus on two influences on discourse acceptability: thematic role bias and Question Under Discussion structure. Both factors are ones that should not, a priori, have anything specifically to do with ellipsis, and indeed both influences on interpretation have been motivated by numerous studies unrelated to ellipsis (Tanenhaus et al. 1994; McRae et al. 1998; Christianson et al. 2001; Roberts 1996; Rohde and Kehler 2009). We show that for both thematic role bias and Question Under Discussion structure, elliptical and non-elliptical sentences are affected to the same extent, unlike structural mismatch effects, which interact with the presence of VPE.

4.1 Experiment 3a: Evidence for a general, strong effect of thematic role bias

To test the degree to which there is a general preference for adjacent clauses to match in voice, and then demonstrate that this general preference is not sufficient to explain the heightened voice sensitivity present in VPE, we examine thematic role bias, which has a broad influence on discourse acceptability. We first use this general bias to calibrate the level of the general preference for matching voice in sentences without ellipsis (Experiment 3a): thematic role bias is in general stronger than the preference for matching voice in non-elliptical sentences. Turning to VPE (Experiment 3b), we show that the structural and thematic role biases reverse under ellipsis: the thematic role

bias is easily overridden by the preference for matching voice, but only in sentences with VPE.

4.1.1 Design

Using a sentence completion task, we pitted voice match against thematic match—that is, having matching thematic role assignments to the subject across clauses. Incomplete sentences like those in (21) were created. The first clause appeared in either active (21a)–(21c) or passive (21b)–(21d) voice. The bias of the final argument (the first argument in the incomplete second clause) was manipulated—it was either agent-biased (21a)–(21b), or patient-biased (21c)–(21d), given the content of the first clause.

(21) a. *Agent-bias, active voice:*
      The medic treated the injured girl in the ambulance, and at the hospital, an ER doctor _____

     b. *Agent-bias, passive voice:*
      The girl who was most badly injured was treated by the medic in the ambulance, and at the hospital, the ER doctor _____

     c. *Patient-bias, active voice:*
      The medic treated the injured girl in the ambulance, and at the hospital, the other children _____

     d. *Patient-bias, passive voice:*
      The girl who was most badly injured was treated by the medic in the ambulance, and at the hospital, the other children _____

In choosing the biasing arguments, we took into account anything that increased the plausibility of an argument in an agent or patient role, including animacy, and lexical material elsewhere in the sentence. The primary determinant of bias was the plausibility of a parallel or contrastive interpretation of the argument in the second (incomplete) clause with one of the arguments in the first clause, together with the goodness of fit between the main verb and the argument in the first clause as either the agent or the patient of that verb [for example, in (21), an ER doctor is a highly plausible parallel or contrastive argument to a medic, and a medic is highly plausible as the agent of the verb 'treat' in the first clause].

In two of the four conditions [(21a), (21d)], following thematic role bias yields the same structure as matching syntactic structure. In the other two conditions [(21b), (21c)], thematic bias conflicts with structural match in terms of the structure predicted.[15] If voice match across clauses is preferred more strongly than thematic role match, we expect more completions which continue the second sentence using the same voice as the first clause; on the other hand, if thematic role match is a more dominant bias, we expect more second sentence completions in which the subject has the same thematic role as the subject of the first sentence—for example, an agent-biased argument should be more likely to elicit an active completion than a passive one.
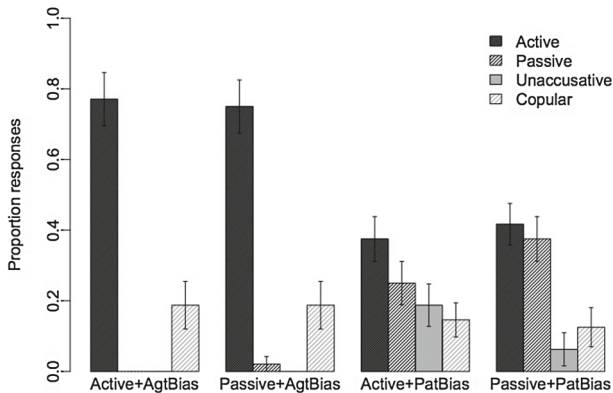
Fig. 3 Experiment 3a proportion completions (*error bars* represent standard error)

4.1.2 Method and procedure

In an experiment run using Mechanical Turk,[16] we asked participants to complete sentence stems like those in (21). 24 native English speakers participated in the study.

4.1.3 Results

Completions were coded as one of the following structures: active voice (including transitive, unergative intransitive, and sentential complement), passive voice, unaccusative, or copular. This covered 96.2% of responses. The results suggest that thematic role bias is a far stronger predictor of the syntactic forms produced by participants than is a bias toward matching syntactic forms. As shown in Fig. 3, sentences with an agent-biased argument were for the most part completed using active voice, regardless of the structure of the preceding clause. Completions after patient-biased arguments were more varied, however there were more passive completions (and fewer active completions) than after agent-biased arguments.

The subset of the data with transitive responses (i.e. responses that were unambiguously actives or passives, corresponding to 74.3% of the original dataset) was analyzed using a mixed-effects logistic regression model predicting response type (passive vs. active voice), with the voice of the first clause, thematic role bias, and their interaction as fixed effects. Only thematic bias is a reliable predictor of completion structure (Table 5). This suggests that, whatever bias there might be to match voice across clauses, it can easily be overridden by other factors—here, the bias to continue with an argument bearing the same thematic role.[17]

---

[16] Amazon's online crowdsourcing platform (http://www.mturk.com).

[17] In fact, as pointed out by a reviewer, it may be that thematic bias or animacy effects would appear even in isolated clauses (e.g. if completion prompts were simply animate or inanimate subject arguments without a preceding clause). This is entirely compatible with the broader point we want to make here: we know that features of lexical items such as (in)animacy and goodness of thematic fit will affect how easy any sentence is to interpret, irrespective of VPE. Acceptability in VPE therefore cannot avoid being affected by such across-the-board biases.

Table 5 Experiment 3a model and model coefficients

```
Response ~ Clause1Passive + Clause2RoleBias
    + Clause1Passive:Clause2RoleBias
    + (1 + Clause1Passive + Clause2RoleBias|Subject)
    + (1 + Clause1Passive + Clause2RoleBias|Item)
```

	β	SE	t	p
Intercept	−0.49	0.20	−2.48	*
Clause1Passive	0.029	0.089	0.33	*n.s.*
Clause2RoleBias	0.48	0.21	2.29	*
Clause1Passive:Clause2RoleBias	−0.0051	0.040	−0.13	*n.s.*

Now that we have shown that thematic bias strongly influences expectations about upcoming structure, we turn back to ellipsis.

4.2 Experiment 3b: Acceptability in ellipsis is modulated by thematic bias

4.2.1 Design

Adapting the materials from Experiment 3a, we created (complete) sentences that varied along three dimensions, resulting in eight conditions [Voice mismatch (Match, Mismatch) × Thematic bias congruence (Bias-congruent, Bias-incongruent; whether the sentence resolved in a manner congruent with thematic role bias) × Ellipsis (Ellipsis, No ellipsis)]. In addition, half of the items had active antecedents and half had passive antecedents. An example is shown in (22)–(23).

(22) a. *Voice match, bias-congruent, ellipsis:*
       The medic treated the injured girl in the ambulance, and at the hospital, an ER doctor did, too.
      b. *Voice match, bias-incongruent, ellipsis:*
       The medic treated the injured girl in the ambulance, and at the hospital, the other children did, too.
      c. *Voice mismatch, bias-congruent, ellipsis:*
       The medic treated the injured girl in the ambulance, and at the hospital, the other children were, too.
      d. *Voice mismatch, bias-incongruent, ellipsis:*
       The medic treated the injured girl in the ambulance, and at the hospital, the ER doctor was, too.

(23) a. *Voice match, bias-congruent, no ellipsis:*
       The medic treated the injured girl in the ambulance, and at the hospital, an ER doctor treated her, too.
      b. *Voice match, bias-incongruent, no ellipsis:*
       The medic treated the injured girl in the ambulance, and at the hospital, the other children treated her, too.

    c. *Voice mismatch, bias-congruent, no ellipsis:*
       The medic treated the injured girl in the ambulance, and at the hospital, the other children were treated by him, too.

    d. *Voice mismatch, bias-incongruent, no ellipsis:*
       The medic treated the injured girl in the ambulance, and at the hospital, the ER doctor was treated by him, too.

4.2.2 Method and procedure

Participants rated the sentences on a 1-to-7 acceptability scale, on Mechanical Turk, with higher ratings corresponding to more acceptable sentences. 16 native English speakers participated in the study.

4.2.3 Results and discussion

We fit the ratings using a mixed-effects regression model, with Bias-congruence, Voice mismatch, and Ellipsis as fixed effects (including all interactions), and the maximal random effects structure justified by the data. The resulting model coefficients are given in Table 6. Condition means are plotted in Fig. 4.

First, both bias incongruence and ellipsis independently decrease acceptability. Turning to the interactions, the Mismatch-Ellipsis interaction replicates the basic results from Experiments 1 and 2: the degradation due to mismatch is stronger when there is ellipsis. By contrast, neither of the interaction terms involving Congruence and Ellipsis reaches full significance, suggesting that thematic bias influences accept-

Table 6 Experiment 3b model and model coefficients

```
log(Acceptability) ~ IncongruentBias + Mismatch + Ellipsis
     + IncongruentBias:Mismatch + IncongruentBias:Ellipsis
     + Mismatch:Ellipsis + IncongruentBias:Mismatch:Ellipsis
     + (1 + IncongruentBias + Mismatch + Ellipsis|Subject)
     + (1 + Mismatch + Ellipsis|Item)
```

	β	*SE*	*t*	*p*
Intercept	3.90	0.34	11.59	***
IncongruentBias	−0.68	0.10	−6.63	***
Mismatch	−0.25	0.12	−2.12	.
Ellipsis	−0.48	0.15	−3.43	**
IncongruentBias:Mismatch	0.20	0.075	2.70	**
IncongruentBias:Ellipsis	0.15	0.075	1.96	.
Mismatch:Ellipsis	−0.33	0.075	−4.44	***
IncongruentBias:Mismatch:Ellipsis	0.14	0.075	1.90	.
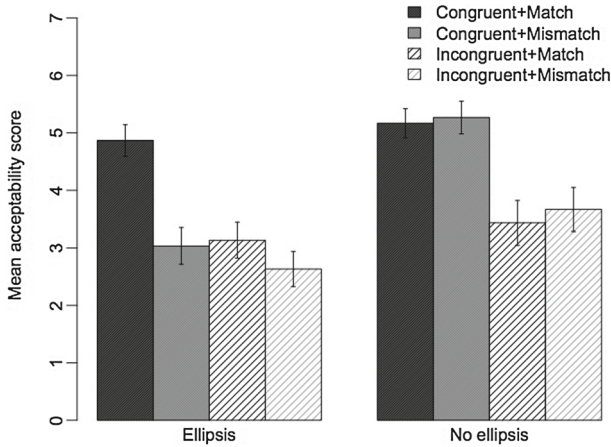
Fig. 4 Experiment 3b mean acceptability ratings (*error bars* represent standard error)

ability in ellipsis because it influences acceptability in general. In addition, there is a Congruence-Mismatch interaction, such that bias-congruent sentences show a smaller mismatch penalty than do bias-incongruent sentences. We speculate that this asymmetry is due to bias-congruent sentences being more easily interpreted than bias-incongruent ones, and therefore more resilient to structural mismatch.

One possible interpretation of these results is that some VPE sentences are ambiguous, and multiple factors, including the structural constraint on VPE and the thematic role bias based on plausibility, collectively determine the interpretation that the parser finally arrives at. In fact, Garnham and Oakhill (1987) have shown that people often do adopt a plausible interpretation instead of a linguistically well-formed one, in cases where these factors are in conflict with each other. In the current experiment, only sentences containing passive voice in the ellipsis clause are potentially ambiguous— examples based on (22c) and (22d) are given in (24), with the intended reading underlined.

(24) a. The medic treated the injured girl in the ambulance, and at the hospital, the other children were {treated, treating the injured girl}, too.
      b. The medic treated the injured girl in the ambulance, and at the hospital, the ER doctor was {treated, treating the injured girl}, too.

In (24a), the intended reading, which has mismatched voice, also has congruent bias; that is, 'the other children' is a better patient than it is an agent for 'treat'. Adopting the intended interpretation therefore requires violating structural identity, but in a way that may be mitigated by plausibility. In (24b), however, the intended reading has mismatched voice and incongruent bias: 'the ER doctor' is a worse patient than it is an agent for the verb. When both structural identity and plausibility are unfavorable, might participants have adopted the alternate reading, which is plausible and does not involve voice mismatch, but instead involves aspectual mismatch between 'treated' in the first clause and 'was treating' in the second?

Table 7 Experiment 3b model and model coefficients—unambiguous items only

```
log(Acceptability) ~ IncongruentBias + Mismatch + Ellipsis
    + IncongruentBias:Mismatch + IncongruentBias:Ellipsis
    + Mismatch:Ellipsis + IncongruentBias:Mismatch:Ellipsis
    + (1 + IncongruentBias + Mismatch + Ellipsis|Subject)
    + (1 + Mismatch + Ellipsis|Item)
```

	β	SE	t	p
Intercept	4.03	0.35	11.65	***
IncongruentBias	−0.76	0.13	−5.68	***
Mismatch	−0.14	0.12	−1.11	*n.s.*
Ellipsis	−0.35	0.14	−2.48	*
IncongruentBias:Mismatch	0.16	0.087	1.86	.
IncongruentBias:Ellipsis	0.076	0.087	0.87	*n.s.*
Mismatch:Ellipsis	−0.22	0.087	−2.52	*
IncongruentBias:Mismatch:Ellipsis	0.10	0.087	1.19	*n.s.*

Whether the alternate reading is favored depends on how tolerant people are of aspectual mismatch relative to voice mismatch or implausibility. Lasnik (1995) marks such examples as ungrammatical (25a), and the unelided version as grammatical (25b).

(25)   a. *John slept, and Mary was too. (*Example 62-a from Lasnik 1995*)
        b. John slept, and Mary was sleeping too. (*Example 62-b from Lasnik 1995*)

However, in order to determine whether our results were affected by such ambiguities, we performed a separate analysis on just the portion of the dataset that did not involve ambiguous sentences (75% of the original dataset). The resulting model and model coefficients are given in Table 7. The new model retains the significant effects from Table 6, except that the Congruence-Mismatch interaction is now marginal rather than fully significant. Importantly, it is still the case that Incongruence affects acceptability across the board, while (structural) Mismatch selectively degrades sentences with VPE.

We therefore take Experiment 3b to illustrate that, while there is no general bias favoring voice match in sentence pairs without ellipsis, when the second clause does contain ellipsis, voice match becomes an overriding bias. This lends support to our claim that there is a VPE-specific structural constraint, and that the resulting heightened sensitivity to voice mismatch in VPE cannot be explained by a broader class of constraints governing discourse well-formedness (and in turn, interpretability). We now turn to a similar general bias at the level of discourse structure.

4.3 Experiment 4: Question Under Discussion and discourse well-formedness

Another factor that strongly influences the well-formedness of a discourse is whether discourse units conform to the question-structure of the discourse it is a part of. Here,

we show that QUD clash and structural mismatch have separate, independent effects on acceptability in discourse. As with the lexical effects discussed above, looking at just elliptical discourses can make these effects look very similar to each other. In fact, some prior accounts have appealed to question-structure as the primary determinant of well-formedness in VPE. For instance, Hendriks (2004)'s analysis requires an elliptical sentence to be in the set of possible answers to an implicit question, where the structure and scope of the question is determined either by contrast or particular discourse relations between the antecedent and ellipsis clauses (see also Winkler 2000, and more recent work by Kehler 2015). However, as we will show, comparison of discourses with and without ellipsis reveal that structural and QUD effects differ in generality: only structural mismatch selectively degrades VPE.

Based on Roberts (1996), Büring (2003) (among others), we operationalize Q-match and Q-clash as in (26).

(26) a. Q-MATCH CONSTRAINT:
          Every clause must have a focused constituent bearing the same thematic role as a focused element in the local QUD.
      b. Q-CLASH:
          Any clause that violates Q-match creates a Q-clash.

The QUD-sentence pairs in (27a)–(27b) and (28a)–(28b) exemplify Q-match and Q-clash, respectively. The (b) sentences use clefting to indicate that the underlined elements are focused.

(27) a. Q: <u>What</u> did Kate order?
      b. A_1: It was <u>a latte</u> that Kate ordered.
      c. A_2: Kate ordered <u>a latte</u> and she ordered <u>a bagel</u>, too.
(28) a. Q: <u>What</u> did Kate order?
      b. A_1: It was <u>Kate</u> who ordered a latte.
      c. A_2: <u>Kate</u> ordered a latte and <u>Ann</u> ordered one, too.

Note that, in the alternative answers in (27c) and (28c), the coordinated clauses keep one argument constant [*Kate/she* in (27c) and *a latte/one* in (28c)] while varying another [*a latte/a bagel* in (27c) and *Kate/Ann* in (28c)]. This serves as an implicit cue to focus—just like the cleft construction in the (b) sentences. (27c) is therefore an instance of Q-match, as an answer to (27a), and (28c) an instance of Q-clash, as an answer to (28a). Experiment 4 makes use of sentences like these to disentangle effects of Q-clash and voice mismatch.

4.3.1 Design

We constructed discourses consisting of a question (representing the QUD) followed by a pair of sentences. The first sentence was an introductory discourse topic. The second sentence—the test sentence—was an elaboration of the discourse topic, and the answer to the question. There were eight conditions in the experiment [Voice

mismatch (Match, Mismatch) × Q-clash (Q-match, Q-clash) × Ellipsis (Ellipsis, No ellipsis)].[18] An example is given in (29)–(32).

In the example in (29), the test sentence both has the same focus structure as the question it is supposed to answer (contrastive arguments underlined), and has conjuncts that match in voice (either active or passive).

(29)   *Q-match, voice match:*
       a.   <u>What</u> was featured on the front page of the Larchmont Chronicle?
       b.   The editorial staff of the Larchmont Chronicle had a hard time deciding which stories should go on the front page.
       c.   <u>The highly publicized murder trial</u> was featured on the front page, and <u>the local business story</u> {was,was featured on the front page}, too.

In the Q-match-Voice mismatch condition, the test sentence conformed to the focus structure of the question, but had clauses that mismatched in voice, as in (30).

(30)   *Q-match, voice mismatch:*
       a.   <u>What</u> was featured on the front page of the Larchmont Chronicle?
       b.   The editorial staff of the Larchmont Chronicle had a hard time deciding which stories should go on the front page.
       c.   The front page featured <u>the highly publicized murder trial</u>, and <u>the local business story</u> {was,was featured on the front page}, too.

Note that it is possible to interpret the first clause of (30c) with a focus structure that clashes with that of the question in (30a) (*The front page featured the highly publicized murder trial…*), since participants did not see any underlining to indicate the constituent meant to be focused. However, because the second clause continues by mentioning 'the local business story', which can be contrasted with 'the highly publicized murder trial' but not 'the front page', (30c) as a whole can only plausibly be interpreted as having a matching focus structure to (30a). Even when the second clause does not contain VPE, the unelided material 'on the front page' cannot plausibly be interpreted as contrasting with any argument from the preceding clause.

The Q-clash-Voice match and Q-clash-Voice mismatch conditions are illustrated in (31)–(32).

(31)   *Q-clash, voice match:*
       a.   <u>What</u> was featured on the front page of the Larchmont Chronicle?
       b.   The editorial staff of the Larchmont Chronicle had a hard time deciding which stories should go on the front page.
       c.   <u>The front page of the Larchmont Chronicle</u> featured the highly publicized murder trial, and <u>the New York Times</u> {did,featured it}, too.

---

(32)   *Q-clash, voice mismatch:*
    a. <u>What</u> was featured on the front page of the Larchmont Chronicle?
    b. The editorial staff of the Larchmont Chronicle had a hard time deciding which stories should go on the front page.
    c. The highly publicized murder trial was featured on <u>the front page of the Larchmont Chronicle</u>, and <u>the New York Times</u> {did, featured it}, too.

The same argument applies to (31c) and (32c) as to (30c): although it would be possible to assign a matching focus structure to the first clause in (31c) or (32c), the material in the second clause, by highlighting the contrast between 'the front page of the Larchmont Chronicle' and 'the New York Times', makes this interpretation highly implausible for the sentence as a whole.[19]

4.3.2 Method and procedure

25 participants on Mechanical Turk rated the test sentences for how natural they sounded as answers to the question, on a 1-to-7 scale, with higher ratings corresponding to more natural-sounding answers.

4.3.3 Results and discussion

The mean ratings by condition are shown in Fig. 5. The data was fit with a mixed-effects regression model including Ellipsis, (Voice) Mismatch, Q-clash, and their interactions as fixed effects. The coefficient estimates are given in Table 8.

There were main effects of Voice mismatch and Q-clash, such that voice-mismatched sentences were rated lower than matched ones, and sentences whose focus structure clashed with that of the QUD were rated lower than those whose focus structure matched the QUD. The Mismatch-Ellipsis interaction replicates the findings of Experiments 1, 2 and 3b: the degradation due to mismatch is larger when the sentence contains ellipsis than when it does not. Again, this suggests that structural match (here represented by voice) is a selective requirement in sentences with VPE.

However, unlike the structural mismatch effect, the degradation associated with QUD clash does not interact with the presence of ellipsis, suggesting that conforming to the focus structure of the QUD is a general constraint on discourses. This is what we would independently expect, since the QUD framework and other discourse structure frameworks are intended to account for general well-formedness in discourse (e.g. Roberts 1996; Büring 2003).

---

[19] There is an additional interpretive possibility in these cases, compared to (29)–(30): the (c) sentences could serve as a cue to a higher level pair-list QUD like *What (story) was featured by which paper?* The *which paper* argument could then be construed as a contrastive topic (cf. Büring 2003), with the *what (story)* argument as the focus. Under such an analysis, the second clause in the elided versions of the (c) sentences triggers a Q-clash because the remnant of ellipsis (*the New York Times*) is not a focus constituent, though it would correspond to a focused element in the QUD. Note that (9), from Kertz (2013), can be analyzed analogously, with a question like *What things are how easy/difficult to identify?* serving as the QUD for *It's easy to identify venomous snakes, and poisonous plants are as well*.
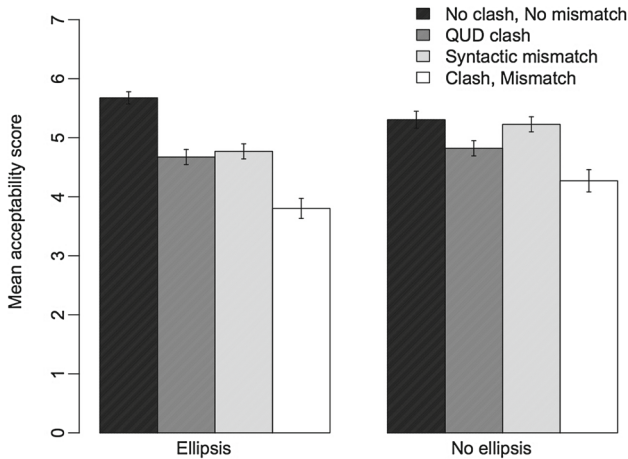
Fig. 5 Experiment 4 mean acceptability ratings (*error bars* represent standard Error)

Table 8 Experiment 4 model and model coefficients

```
log(Acceptability) ~ Ellipsis + Mismatch + Q-clash
     + Ellipsis:Mismatch + Ellipsis:Q-clash + Mismatch:Q-clash
     + Ellipsis:Mismatch:Q-clash + (1 + Ellipsis + Mismatch
     + Q-clash|Item) + (1 + Ellipsis + Mismatch + Q-clash|Subject)
```

	β	SE	t	p
Intercept	4.89	0.13	38.28	***
Ellipsis	−0.11	0.12	−0.95	*n.s.*
Mismatch	−0.40	0.075	−5.37	***
Q-clash	−0.46	0.096	−4.77	***
Ellipsis:Mismatch	−0.23	0.080	−2.82	*
Ellipsis:Q-clash	−0.013	0.094	−0.14	*n.s.*
Mismatch:Q-clash	−0.016	0.082	−1.93	.
Ellipsis:Mismatch:Q-clash	0.051	0.092	0.55	*n.s.*

The approach taken by Kehler (2015) characterizes acceptable VPE in terms of a QUD antecedent—a clause containing VPE must be in the alternative set denoted by the QUD. In this respect, Kehler's proposal is like others (e.g. Dalrymple et al. 1991; Hardt 1993; Hendriks 2004; Kertz 2008, 2010) that seek to explain VPE without any ellipsis-specific constraints, by appealing to broader constraints on discourse well-formedness. Of particular relevance here is that these proposals do not include any means for distinguishing VPE which structurally matches its antecedent from VPE which does not. While we agree that the grounding of such approaches in independently-motivated discourse well-formedness constraints is appealing, our findings show that voice mismatch degrades VPE acceptability beyond what can be

explained by QUD constraints. Much like we saw in Experiment 3b, Experiment 4 shows that structural match is a constraint isolated to clauses containing VPE, while the QUD effect is a general constraint on discourses. Thus, QUD-related well-formedness affects sentences across the board, while violations of structural parallelism selectively degrade VPE sentences. This suggests that a constraint explicitly ruling out structural mismatch is needed to explain the distribution of acceptability in VPE.

5 General discussion

It is apparent that the perceived acceptability of sentences with VPE is influenced by a combination of lexical and discourse-level factors—including the coherence relation linking the antecedent and ellipsis clauses, thematic role bias, and QUD well-formedness—which affect interpretability. However, in the current study we have shown that there is a systematic structural mismatch penalty in VPE which persists even when these other factors have been controlled. We take this to indicate that the primary mechanism for resolving VPE does engage structural representations.

In the remainder of this section, we discuss some implications of our findings, and questions that remain open. First, we spell out our view on the relationship between grammaticality and interpretability. Next, we consider the possibility that syntactic structure is a part of discourse structure, rather than a separate level of representation. Finally, we touch on some relevant questions that are not addressed by this paper, which we leave for future study.

5.1 Grammaticality, acceptability and interpretability

The results from Experiments 1–4 are consistent with and build upon previous psycholinguistic and theoretical work (Tanenhaus and Carlson 1990; Fiengo and May 1994; Johnson 2001; Kennedy 2003; Arregui et al. 2006, among others) all pointing to the existence of a structural licensing condition on VPE. But even if we accept that there is a structural identity condition on VPE, there is the further question of how to account for cases when there fails to be perfect structural match between the antecedent and elided VPs. Since we observe that even in the face of structural mismatch, comprehenders are able to understand what a sentence means, there must be a way to parse and interpret such sentences. On the production side as well, a speaker may produce an instance of mismatched VPE despite its violation of structural identity, because various processing pressures can conflict either with each other or with grammatical constraints. For example, Bock and Miller (1991) showed that speakers often produce ungrammatical agreement errors when a noun with different number intervenes between the head noun and the agreeing verb (e.g. *The cost of the improvements have not yet been estimated*).

This raises a set of questions about the relationships among grammaticality, acceptability, and interpretation. It is important to acknowledge that the distinctions among these notions are largely a matter of the kind of theory a researcher wants to construct; that is, they are not (so far) empirically decidable by us or by any other researchers who have taken stands on this issue. What we do know is that the grammar (linguis-

tic competence, in its broadest construal) is embedded in linguistic behavior, which means that we cannot study things about the grammar either directly or in isolation— we must always study them through the filter of linguistic behavior.[20] Judgments of acceptability (the primary measure we deal with) are linguistic behavior. As such, explanations of patterns of acceptability can and do include grammatical explanations, as well as ones invoking processes and mechanisms associated with comprehending or producing language (e.g. lexical retrieval mechanisms, accessing conceptual knowledge networks) which may not even be inherently linguistic (e.g. inferences based on information from the visual context informing syntactic parses). When we invoke the notion of interpretability, we are advancing a hypothesis: that the acceptability patterns we observe are at least in part interpretability patterns—changes in acceptability due to difficulty or ease of interpretation. [See Francom (2009) for related work linking interpretability to acceptability.]

Based on our experiments involving voice mismatch, we have proposed that there is a grammatical constraint on VPE requiring that the elided VP and its antecedent have structurally matching representations. Two additional points are crucial to our proposal. Firstly, imperfectly matching instances of VPE which violate this structural constraint can still receive an interpretation. Secondly, we believe that acceptability in VPE is ultimately about interpretability: violations of the structural constraint result in degraded acceptability, but *less so* if other (e.g. inferential, pragmatic) means are available to help the comprehender assign an interpretation. (33) provides some cases of mismatched but acceptable-sounding VPE (examples are from Webber 1978).

(33) a. China is a country that Joe wants to visit, and he will too, if he gets an invitation there soon.
      b. Martha and Irv wanted to dance together, but Martha couldn't, because her husband was there.
      c. Wendy is eager to sail around the world and Bruce is eager to climb Kilamanjaro, but neither of them can because money is too tight.

We suspect such cases are precisely where inferential mechanisms may suggest interpretations that would otherwise be harder to access given the structural mismatch.[21] While we have only considered sentences with voice mismatch in the current study, an important goal moving forward will be to extend the current findings to more varied forms. The decision to focus on the voice alternation in this set of experiments was in large part tactical. Structural mismatch effects are reported in some form or other by a number of researchers investigating VPE, in a range of constructions (Tanenhaus and Carlson 1990; Arregui et al. 2006; Frazier and Clifton 2006; Fine et al. 2009; Kertz 2010, 2013; Kim et al. 2011); we therefore wanted to take the first step of investigat-

---

[20] As stated in Bever (1970), the following (often misquoted to mean the opposite of what it means):

(100) (Apparent Linguistic Universals) − (Cognitive Universals) = Real Linguistic Universals

is precisely something we *cannot* do: "such an enterprise fails to take into consideration the fact that the influences of language and cognition are mutual; one cannot consider one without the other" (p. 352).

[21] Thanks to a reviewer for pointing out these examples.

ing the properties of this structural mismatch effect, using a structural alternation that has been very well-studied in the theoretical literature, including specifically in the domain of ellipsis. Moving forward, it will be of great interest to narrow down precisely what structural properties are relevant for the purpose of interpreting VPE, and thereby better understand how the structural restriction interacts with other linguistic pressures to collectively determine acceptability.

Having a structural constraint that can be satisfied to varying extents also allows for *degrees* of acceptability. Prior research (Arregui et al. 2006; Kim et al. 2011) has argued for models of VPE resolution that link the extent of mismatch to the extent of degradation in acceptability. In terms of the relationship between grammar and acceptability, Kim et al. (2011) (and Kobele et al. 2008) assumed that cases of structural mismatch are grammatical in that they were generated by the grammar. Instead, they linked degradations in acceptability to the size of the elided constituent in the derivation tree. Structural mismatch will generally have the consequence that a smaller subpart of the antecedent and elided VP structures will be identical; this, combined with a parsing preference in the spirit of MaxElide (cf. Merchant 2008; Takahashi and Fox 2005) that prefers larger constituents to be elided, was argued to predict the graded pattern of acceptability observed, with greater degrees of mismatch yielding greater degradation.[22]

An advantage to the view just described is that structurally non-parallel structures can be generated in a normal way, and as such do not have to invoke anything outside of the usual mechanisms for computing meanings from syntactic structures. However, proposals have been made which argue for a different relationship between grammaticality and the source of unacceptability. For instance, according to the Recycling Hypothesis proposed in Arregui et al. (2006), cases of structural mismatch are ungrammatical (i.e. the grammar does not generate such cases), and the degradation comes about when a comprehender attempts to assign such ungrammatical sentences an interpretation, which involves engaging a repair process that alters the mismatched structure until it matches.

As mentioned above, it seems to us that no direct empirical arguments exist at this time to definitively favor one model or another (a similar point is made by Phillips and Parker 2014). However, it may be possible, moving forward, to assess whether certain models are more or less likely given the kinds of adjustments they would have to support in order to expand their coverage of data as it becomes available. For instance, Kim et al. (2011)'s model could be extended to explain discourse coherence effects by allowing the discourse context that an instance of VPE appears in to influence the strength of the ellipsis size constraint. By contrast, the model in Arregui et al. (2006) would either have to alter the number of repair steps needed to restore structural identity under certain discourse conditions, allow for repair steps to have differential effects on overall acceptability under certain conditions, or relax the requirement for

---

[22] It should be noted that MaxElide, which may make predictions about the choice of an antecedent among multiple candidate antecedents, and the current proposal, which is about the evaluation of representational match, are not direct alternatives to each other.

strict syntactic identity under certain conditions. We leave this set of questions for future research.[23]

5.2 The right level of structural representation

A number of existing accounts of VPE (e.g. Dalrymple et al. 1991; Hardt 1993; Merchant 2001, 2013; Jacobson 2014) do not predict any degradation related to voice mismatch to begin with, since the antecedent-VPE relationship, or the identity constraint that governs it, operates at a level of representation that does not distinguish between active and passive voice. Our findings argue against such accounts. However, the current study raises a question about the extent to which syntactic and discourse representations should be distinguished, or incorporated into a unified representation. While we do need there to be a specific structural constraint on VPE resolution, we currently remain agnostic as to the exact level of representation at which the constraint applies. We do not, for example, find support for different sets of rules governing ellipsis within and across sentences, as Frazier and Clifton (2005) propose (see Experiment 2, Sect. 3). Can we then consider syntax and discourse as simply different grains of representation, rather than distinct levels?

A possible alternative explanation of the mismatch effects reported here that does not involve a syntactic identity condition is that what we are calling structural mismatch effects are really discourse-level mismatch effects (such a system is adopted by e.g. Hardt and Romero 2004). Since changes to syntactic structure (active vs. passive, nominal or adjectival vs. verbal) are likely to have corresponding effects on discourses (by means of changes in information structure, or predication structure), any of the structural mismatch effects reported here can be interpreted as discourse effects which are only indirectly related to syntactic structure. In fact, such an alternative is considered by Tanenhaus and Carlson (1990) from the point of view of comparing deep and surface anaphora.

It is difficult to pull apart effects of syntactic and discourse structure when they co-occur as they do here, but we note that whatever level of structure is responsible for the sensitivity to voice (or category) mismatches in VPE, it must be at least structurally rich enough that actives and passives (or nominalizations, deverbal adjectives, and verb phrases) have distinct representations. What we have shown here is that the level of representation relevant for VPE resolution must both encode enough structural detail to distinguish active from passive voice (as in Sag and Hankamer 1984), and represent clauses spanning discourse (adjacent sentences) and clauses within a (syntactically connected) sentence in a comparable way.

Relatedly, the combined empirical evidence to date does not clearly distinguish effects of structural match from contrast alignment effects (Winkler 2000; Hendriks 2004; Kertz 2013) (see also Ginzburg and Sag 2000). While we dispute the claim that such information structural effects are able to account for VPE without any ellipsis-

---

[23] See Lau et al. (2014) for a discussion relating gradient acceptability judgments to gradient grammaticality, and Schütze and Sprouse (2014) for discussion of why gradient acceptability is not necessarily incompatible with categorical grammaticality.

specific constraints, it may be that the ellipsis-specific structural constraint we propose is actually an ellipsis-specific constraint enforcing focus structure alignment. An example of such a constraint is the condition on VPE proposed in Rooth (1992), which requires a constituent containing VPE to focus-match an antecedent constituent—roughly, the alternative propositions generated by substituting for focused elements in the ellipsis clause must contain the antecedent proposition. Büring (2005) discusses and extends this proposal to account for the range of strict and sloppy identity interpretations in VPE. Note that the focus-matching requirement builds in a requirement for structural identity between the antecedent and ellipsis clauses.

The work presented here leaves open a number of questions about the relationship between syntactic structure and focus structure. In particular, Experiments 1 and 2 used the Voice alternation to manipulate the syntactic structure of antecedent and elided VPs, but actives and passives differ in a number of ways, including information or focus structure (Vallduví 1992; Birner and Ward 1998) and predication structure (Williams 1980; den Dikken 2006). Here, we have focused our attention on the question of whether there is evidence for a structural parallelism requirement that is specific to VPE—that is, not derivable from general conditions on discourse well-formedness. We leave to future research the interesting question of what the relationship is between general discourse pressures and narrower constraints specific to particular constructions.

6 Conclusion

In the current study, we considered the possibility that an array of discourse and/or lexical biases might be able to explain the pattern of acceptability observed in sentences with mismatched VPE; in this sense, the study can be seen as contributing to recent debates in the literature about whether phenomena traditionally considered grammatical, such as island effects, should be explained in terms of processing biases (e.g. Hofmeister and Sag 2010; Phillips 2012). Indeed, as we noted in the Introduction, we think the goal of explaining the acceptability pattern in terms of general well-formedness biases is a worthwhile one, because the theory would be less stipulative and more explanatory. However, as these experiments show, there continues to be a residual mismatch effect above and beyond the effects of general biases relating to the likelihood or fit of combinations of lexical items, and how a sentence is integrated into the larger discourse structure in a coherent way.

While we remain open to alternative explanations, then, we take our findings to suggest that there is a structural identity constraint on VPE resolution. We propose that the variability in the data is accounted for by a combination of these general biases and the ellipsis-specific structural constraint. Whether this account can be extended to other varieties of ellipsis is a question we leave open for future research. Finally, the current study serves as another demonstration that, because grammaticality must always be studied by way of acceptability (linguistic behavior), effects related to grammatical constraints and those related to complexity of interpretation must also always be carefully considered in relation to each other.

Appendix A: Experimental 1 items

For each list, each item appeared in one of the four following conditions:

- Match
  - Ellipsis
  - No ellipsis
- Mismatch
  - Ellipsis
  - No ellipsis

Half of the items appeared with active voice in the first clause, and half with passive voice in the first clause.

RESEMBLANCE *items*

1. Almost everyone accused Bill of stealing the money, but the secretary didn't./ didn't accuse him.
   Everyone accused Bill of stealing the money, but the real culprit wasn't./wasn't accused by anyone.
2. The paramedics examined the survivors of the plane crash, and the psychologist did, too./examined them, too.
   The psychologist examined the survivors of the plane crash, and their family members were, too./were examined, too.
3. Bruce burned himself on the stove, and Jason did, too./burned himself, too.
   Bruce burned himself on the stove, but Jason wasn't./wasn't burned.
4. The play was advertized by the university, but the concert wasn't./wasn't advertized.
   The play was advertized by the students, but the university didn't./didn't advertize it.
5. Lisa was suspected by the cops, and her boyfriend was, too./was suspected, too.
   Lisa was suspected by the cops, and the private investigator did, too./suspected her, too.
6. Sara was voted off the team, and Ellen was, too./was voted off, too.
   Sara was voted off the team by the boys, but the girls didn't./didn't vote her off.
7. Sam recommended the New York Times bestseller to Meg, and Todd did, too./recommended it to her, too.
   Someone recommended the New York Times bestseller to Meg, and the new Kennedy biography was, too./was recommended to her, too.
8. Amy introduced herself to the guest speaker, and Jane did, too./introduced herself, too.
   Amy introduced herself to the guest speaker, and Jane was, too./was introduced to him, too.

9. John reported the vandalism at the office, and Lana did, too./reported it, too.
   Someone reported the vandalism at the office, but the stolen files weren't./weren't reported.

10. Pete was nominated for the award by his advisor, and his officemate was, too./was nominated, too.
    Pete was nominated for the award by his advisor, and his officemate did, too./nominated him, too.

11. The mayor was interviewed by the local paper, and the sheriff was, too./was interviewed, too.
    The mayor was interviewed by the local paper, and CNN did, too./interviewed him, too.

12. The movie was criticized by reviewers, and the new album was, too./was criticized, too.
    The movie was criticized by reviewers, and Oprah did, too./criticized it, too.

13. Lauren passed the midterm, but Nate didn't./didn't pass.
    Everyone passed the midterm, but the final wasn't./wasn't passed by everyone.

14. Molly arranged an interview with the minister, and Eric did, too./arranged one, too.
    Some journalists arranged private interviews with the minister, but a press conference wasn't./wasn't arranged.

15. Alice taped the lecture, and Robin did, too./taped it, too.
    Some students taped the lecture, but the review session wasn't./wasn't taped by anyone.

16. The essay was copied by the student, but the term paper wasn't./wasn't copied.
    The essay was copied by some students, but Jim didn't./didn't copy it.

17. The garage door was repaired over the weekend, and the broken swing set was, too./was repaired, too.
    The garage door was repaired over the weekend, but Tom didn't./didn't repair it.

18. The writer's strike was covered by the media, but the teacher's strike wasn't./wasn't covered.
    The writer's strike was covered by the national media, but local papers didn't./didn't cover it.

19. The crime lab analyzed the blood sample, and the private clinic did, too./analyzed it, too.
    The crime lab analyzed the blood sample, and the fingerprints were, too./were analyzed, too.

20. Jill scheduled a meeting with the dean, and Neal did, too./scheduled one, too.
    The faculty scheduled a meeting with the dean, and an admissions committee meeting was, too./was scheduled, too.

21. Kyle gave some advice to the new students, and Katie did, too./gave them advice, too.
    Everyone gave the new students advice, and the graduates were, too./were given advice, too.

22. Tim's proposal was praised by the committee, and Annie's was, too./was praised, too.

Tim's proposal was praised by the committee, and his sponsors did, too./praised it, too.

23. The paper was approved by the editorial board, and the book review was, too./was approved, too.
    The paper was approved by the editorial board, and the anonymous reviewers did, too./approved it, too.
24. Matt was picked up at the airport, but Jim wasn't./wasn't picked up.
    Matt was picked up at the airport, but the department didn't./didn't pick him up.

CAUSE-EFFECT *items*

25. Abby insisted that Bill get rid of the video tape, so he did./destroyed it.
    Abby insisted that Bill's video tape be destroyed, so he did./got rid of it.
26. Greta begged her brother to kill the spider, so he did./killed it with his shoe.
    Greta begged anyone to kill the spider, so it was./was killed by her roommate's boyfriend.
27. There was a consensus that Jack should teach the seminar, so he did./taught it.
    There was a consensus that someone new should teach the seminar, so it was./was taught by Jack.
28. Roy thought the dog should be buried in the backyard, so it was./was buried there.
    Roy thought the dog should be buried in the backyard, so they did./buried it there.
29. Nina demanded that the project be evaluated by the review board, so it was./was discussed at the next meeting.
    Nina demanded that the project be evaluated by the review board, so they did./discussed it at the next meeting.
30. Frank said that the computer needed to be replaced, so it was./was replaced with a better one.
    Frank said that the computer needed to be replaced, so they did./replaced it with a better one.
31. Dana had recommended that Meg take the class, so she did./took it.
    The faculty had recommended that the department cancel the class, so it was./was cancelled.
32. Sue had requested that the presenters dim the lights during the announcement, so they did./turned the lights down.
    Sue had requested that the presenters dim the lights during the announcement, so they were./were turned down.
33. Tom insisted that the committee recount the ballots, so they did./recounted them.
    Tom insisted that the committee recount the ballots, so they were./were recounted.
34. Everyone wanted Jeff to be nominated, but he wasn't./wasn't nominated by anyone.
    Everyone wanted Jeff to be nominated, but no one did./nominated him.
35. It was obvious that the rules had to be amended, so they were./were revised over the summer.
    It was obvious that the rules had to be amended by the group members, so they did./some group members revised them over the summer.
36. The parents thought the children should be examined by a pediatrician, so they were./were all examined.

The parents thought the children should be examined by the school nurse, so she did./examined them all.

37. It was agreed that Mike should sell the car, so he did./sold it on Craigslist.
It was agreed that someone should sell the car, so it was./was sold on Craigslist.

38. Few parents wanted to chaperone the party, so no one did./chaperoned it.
Few parents wanted to chaperone the party, so it wasn't./wasn't chaperoned.

39. Andrea asked Jim to turn on the heat in their apartment, so he did./turned it on when he got home.
Andrea asked Jim to turn on the heat in their apartment, so it was./was turned on when she got home.

40. Most people agreed that the merger should be avoided by the company, so it was./was avoided for a long time.
Most people agreed that the merger should be avoided by the company, so they did./avoided it for a long time.

41. Justin recommended to the judges that the prize be given to the freshmen, so it was./was given to them.
Justin recommended to the judges that the prize be given to the freshmen, so they did./gave it to them.

42. There was a consensus that the earnings should be split equally, so they were./were divided amongst the participants.
There was a consensus that the earnings should be split equally amongst the participants, so they did./divided them amongst themselves.

43. Greg begged Sue to cover his shift on Friday, so she did./covered it.
Greg needed someone to cover his shift on Friday, so it was./was covered for him.

44. Emily's dad told her to forge his signature, so she did./forged it on the letter.
Emily told people to forge her signature on the letter, so it was./was forged.

45. Hardly anyone wanted to revise the bylaws, so no one did./revised them this year.
Hardly anyone wanted to revise the bylaws, so they weren't./weren't revised this year.

46. Everyone agreed that the building should be demolished, so it was./was torn down.
Everyone agreed that the building should be demolished, so they did./tore it down.

47. The class requested that the exam be rescheduled, so it was./was moved to a different time.
The class requested that the exam be rescheduled by the professor, so she did./moved it to another time.

48. Max asked his roommates to water his plant while he was away, so they did./Max asked his roommates to take care of his plant, so they watered it while he was away.
Max asked his roommates to water his plant while he was away, so it was./Max asked his roommates to take care of his plant, so it was watered while he was away.

Filler items

The fillers for Experiment 1 were all grammatical, and varied in length (representative examples given below). Sentence lengths were varied by including additional prepositional phrases or embedded clauses.

- Sandy read a book about the Civil War for her history class.
- Jack indicated that no one should wake him up in the morning.
- The highly anticipated movie disappointed everyone who watched it on opening night.
- Matt wondered whether his roommate had meant to leave his light on all weekend.
- Dan's mother made him buy some new shoes before his big job interview with the law firm.

Appendix B: Magnitude estimation instructions

Participants in Experiments 1 and 2 saw the following instructions.
"In this experiment, you're going to judge the acceptability of some English sentences. Your task will be to judge how good or bad each sentence sounds by assigning a number to it.
First, we're going to practice this by estimating the lengths of lines relative to each other."

[examples using line lengths]

"Now you're going to do the same thing with a bunch of lines of different lengths, except that you get to choose the first line length that you'll compare all the other lengths to.

Whatever length you say the first line is, you should estimate for each of the following lines how long they are compared to the first length you gave.

You can use any range of positive numbers that you like, including decimal numbers. There is no upper or lower limit to the numbers you can use, except that you cannot use zero or negative numbers."

[practice with line lengths]

"Now you're going to do the same thing, but with sentences. The values you give should indicate how good or natural the sentence sounds."

[examples using sentences]

"Now you're going to get a few sentences to try. Remember, you're giving a score for how good or natural a sentence sounds as a sentence of English—similar to coming up with a value to estimate line lengths.

Like before, first you'll choose a numerical value for how good the first sentence sounds. Then for every sentence after that, you'll assign it a number for how good it sounds compared to that first sentence. If a sentence sounds twice as good as the first sentence, it would get a score that's two times the number you gave the first sentence. If a sentence sounds three times worse than the first sentence, it would get a score that's a third the first sentence's value.

You can use any range of positive numbers that you like, including decimal numbers. There is no upper or lower limit to the numbers you can use, except that you cannot use zero or negative numbers."

[practice with sentences]

"That's it for practice. If you have any questions, ask now. The next screen will start the experiment."

[Choosing the modulus:]

"Give this sentence a number. Then give each sentence after it a number based on how good it sounds compared to the first sentence.

The kids were amused by the cartoon, but their parents weren't."

References

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*, 232–246.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, MA: Cambridge University Press.

Bard, E. G., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, *72*(1), 32–68.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York: Wiley.

Birner, B. J., & Ward, G. (1998). *Information status and noncanonical word order in English*. Amsterdam/Philadelphia: John Benjamins.

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*, 177–192.

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, *23*, 45–93.

Bonatti, L. (2008). *PsyScope X, Build 53*. PsyScope X Project. SISSA Language, Cognition and Development Lab. http://psy.ck.sissa.it/.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*, B13–B25.

Büring, D. (2003). On d-trees, beans, and b-accents. *Linguistics & Philosophy*, *26*(5), 511–545.

Büring, D. (2005). *Binding theory*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Cai, Z. G., Pickering, M. J., & Sturt, P. (2012). Processing verb-phrase ellipsis in Mandarin Chinese: Evidence against the syntactic account. *Language and Cognitive Processes*, *28*(6), 810–828.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368–407.

Dalrymple, M., Shieber, S. M., & Pereira, F. C. N. (1991). Ellipsis and higher-order unification. *Linguistics and Philosophy*, *14*, 399–452.

den Dikken, M. (2006). *Relators and linkers: The syntax of predication, predicate inversion, and copulas*. Linguistic Inquiry Monographs. Cambridge: MIT Press.

Fiengo, R., & May, R. (1994). *Indices and identity*. Cambridge: MIT Press.

Fine, A. B., Kim, C. S., & Runner, J. T. (2009). *The role of inflectional entropy in VP ellipsis resolution*. Poster presented at CUNY Conference on Human Sentence Processing. UC Davis.

Francom, J. C. (2009). *Experimental syntax: Exploring the effect of repeated exposure to anomalous syntactic structure—Evidence from rating and reading tasks*, Ph.D. thesis. University of Arizona.

Frazier, L., & Clifton, C. (2005). The syntax-discourse divide: Processing ellipsis. *Syntax*, *8*(2), 121–174.

Frazier, L., & Clifton, C. (2006). Ellipsis and discourse coherence. *Linguistics and Philosophy*, *29*, 319–346.

Garnham, A., & Oakhill, J. (1987). Interpreting elliptical verb phrases. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *39*(4), 611–627.

Ginzburg, J. (2012). *The interactive stance*. Oxford: Oxford University Press.

Ginzburg, J., & Sag, I. (2000). *English interrogative constructions*. Stanford, CA: CSLI Publications.

Grant, M, Jr., Charles, C., & Frazier, L. (2012). The role of non-actuality implicatures in processing elided constituents. *Journal of Memory and Language*, *66*, 326–343.

Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research*, *34*(4), 365–399.

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

Hardt, D. (1993). *Verb phrase ellipsis: Form, meaning, and processing*, Ph.D. thesis. University of Pennsylvania.

Hardt, D. (1999). Dynamic interpretation of verb phrase ellipsis. *Linguistics and Philosophy*, *22*, 187–221.

Hardt, D., & Romero, M. (2004). Ellipsis and the structure of discourse. *Journal of Semantics*, *21*(4), 375–414.

Hendriks, P. (2004). Coherence relations, ellipsis, and contrastive topics. *Journal of Semantics*, *21*(2), 133–153.

Hofmeister, P., & Sag, I. A. (2010). Cognitive constraints and island effects. *Language*, *86*(2), 366–415.

Hofmeister, P., Arnon, I., Jaeger, T. F., Sag, I. A., & Snider, N. (2013). The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes*, *28*(1), 48–87.

Jacobson, P. (2014). Silent linguistic material: An urban legend? In *Proceedings of the 50th annual meeting of the Chicago Linguistic Society*. Chicago.

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). Oxford: Blackwell.

Kaschak, M. P. (2007). Long-term structural priming affects subsequent patterns of language production. *Memory and Cognition*, *35*, 925–937.

Katz, J. J., & Postal, P. (1964). *An integrated theory of linguistic descriptions*. Cambridge, MA: MIT Press.

Kehler, A. (2000). Coherence and the resolution of ellipsis. *Linguistics and Philosophy*, *23*(6), 533–575.

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

Kehler, A. (2015). On QUD-based licensing of strict and sloppy ambiguities. In *Proceedings of semantics and linguistic theory* (Vol. 25).

Kennedy, C. (2003). Ellipsis and syntactic representation. In *The syntax-semantics interface: Interpreting (omitted) structure*. Amsterdam: John Benjamins.

Kennedy, C., & Merchant, J. (2000). Attributive comparative deletion. *Natural Language and Linguistic Theory*, *18*(1), 89–146.

Kertz, L. (2008). Focus structure and acceptability in verb phrase ellipsis. In *Proceedings of the West Coast conference on formal linguistics* (Vol. 27, pp. 283–291). Los Angeles.

Kertz, L. (2010). *Ellipsis revisited*, Ph.D. thesis. UCSD.

Kertz, L. (2013). Verb phrase ellipsis: The view from information structure. *Language*, *89*(3), 390–428.

Kim, C., Kobele, G., Runner, J., & Hale, J. (2011). The acceptability cline in verb phrase ellipsis. *Syntax*, *14*(4), 318–354.

Kobele, G., Kim, C., Hale, J., & Runner, J. (2008). A processing model of ungrammatical VP ellipsis. In *Talk presented at 21st CUNY conference on human sentence processing*. Chapel Hill.

Lasnik, H. (1995). Verbal morphology: Syntactic structures meets the minimalist program. In H. Campos & P. Kempchinsky (Eds.), *Evolution and revolution in linguistic theory: Essays in Honor of Carlos Otero* (pp. 251–275). Washington, D.C.: Georgetown University Press.

Lau, J. H., Clark, A., & Lappin, S. (2014). Measuring gradience in speakers' grammaticality judgements. In *36th Annual conference of the cognitive science society* (pp. 821–826). Quebec City.

Martin, A. E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language*, *58*, 879–906.

Mauner, G., Tanenhaus, M. K., & Carlson, G. N. (1995). A note on parallelism effects in processing deep and surface verb-phrase anaphors. *Language and Cognitive Processes*, *10*, 1–12.

McRae, K., Spivey-Knowlton, M., & Tanenhaus, M. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*, 283–312.

Merchant, J. (2001). *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford: Oxford University Press.

Merchant, J. (2008). Variable island repair under ellipsis. In K. Johnson (Ed.), *Topics in ellipsis* (pp. 132–153). Cambridge: Cambridge University Press.

Merchant, J. (2013). Voice and ellipsis. *Linguistic Inquiry*, *44*(1), 77–108.

Miller, P., & Pullum, G. K. (2013). Exophoric VP ellipsis. In P. Hofmeister & E. Norcliffe (Eds.), *The core and the periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag* (pp. 5–32). Stanford, CA: CSLI Publications.

Miltsakaki, E. (2002). Toward an aposynthesis of topic continuity and intrasentential anaphora. *Computational Linguistics*, *28*(3), 319–355.

Phillips, C. (2012). On the nature of island constraints I: Language processing and reductionist accounts. In J. Sprouse & N. Hornstein (Eds.), *Experimental syntax and island effects* (pp. 64–108). Cambridge: Cambridge University Press.

Phillips, C., & Parker, D. (2014). The psycholinguistics of ellipsis. *Lingua*, *151*, 78–95.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The Penn discourse treebank 2.0. In *Proceedings of the 6th international conference on language resources and evaluation*. Marrakech.

Roberts, C. (1996). Towards an integrated formal theory of pragmatics. In J. H. Yoon & A. Kathol (Eds.), *OSU working papers in linguistics* (Vol. 49, pp. 91–136). Columbus: The OSU Department of Linguistics.

Rohde, H., & Kehler, A. (2009). QUD-driven expectations in discourse interpretation. In *83rd annual meeting of the linguistics society of America*, San Francisco, CA.

Rohde, H., Levy, R., & Kehler, A. (2011). Anticipating explanations in relative clause processing. *Cognition*, *118*(3), 339–358.

Rooth, M. (1992). *Proceedings of the Stuttgart workshop on ellipsis, volume 29 of Arbeitspapiere des SFB 340*, chapter Reduction redundancy and ellipsis redundancy. University of Stuttgart.

Sag, I. (1976). *Deletion and logical form*, Ph.D. thesis. Cambridge, MA: MIT.

Sag, I., & Hankamer, J. (1984). Toward a theory of anaphoric processing. *Linguistics and Philosophy*, *7*, 325–345.

Schütze, C. T., & Sprouse, J. (2014). Chapter 3: Judgment data. In D. Sharma & R. Podesva (Eds.), *Research methods in linguistics*. Cambridge: Cambridge University Press.

Shieber, S. M., Pereira, F., & Dalrymple, M. (1996). Interactions of scope and ellipsis. *Linguistics and Philosophy*, *19*, 527–552.

Sprouse, J. (2011). A test of the cognitive assumptions of magnitude estimation: Commutativity does not hold for acceptability judgments. *Language*, *87*(2), 274–288.

Takahashi, S., & Fox, D. (2005). MaxElide and the re-binding problem. In E. Georgala & J. Howell (Eds.), *Proceedings of SALT 15* (pp. 223–240). Ithaca, NY: CLC Publications.

Tanenhaus, M., & Carlson, G. (1990). Comprehension of deep and surface verb phrase anaphors. *Language and Cognitive Processess*, *5*, 257–280.

Tanenhaus, M. K., Trueswell, J. C., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285–318.

Vallduví, E. (1992). *The informational component*. New York: Garland.

Webber, B. L. (1978). *A formal approach to discourse anaphora*. Technical Report 3761. Cambridge, MA: Bolt, Berenek and Newman.

Wescott, T., & Fanselow, G. (2011). On the informativity of different measures of linguistic acceptability. *Language*, *87*(2), 249–273.

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

Williams, E. (1980). Predication. *Linguistic Inquiry*, *11*, 203–238.

Winkler, S. (2000). Silent copy and polarity focus in VP ellipsis. In K. Schwabe & N. Zhang (Eds.), *Ellipsis in coordination* (pp. 221–247). Tübingen: Niemeyer.