At the syntax-discourse interface: Verb phrase ellipsis interpretation in context

Jeffrey Geiger
University of Chicago

jeffrey.geiger@pomona.edu
185 East Sixth Street
Pomona College LGCS, Edmunds 206
Claremont, CA 91711

Ming Xiang
University of Chicago

mxiang@uchicago.edu
1115 East 58th Street
Rosenwald Hall 205B
Chicago, IL 60637

At the syntax-discourse interface: Verb phrase ellipsis interpretation in context

A central question of language comprehension concerns the interaction between linguistic form and broader representations of discourse in the interpretation of context-sensitive expressions. This interaction is instantiated in the interpretation of verb phrase ellipsis, where previous work has shown that the linguistic antecedent and the broader context are both considered in resolution. Using a novel experimental paradigm, we investigated VPE interpretation in discourses where the antecedent and the broader context make different information available for inclusion in the interpretation of the ellipsis site. Our results point to a complex interaction between linguistic antecedents and the broader discourse context in interpretation, putting considerable constraints on the set of possible models for VPE resolution. This work contributes to a better understanding of both the connections between and the boundaries separating linguistic structure and mental models of discourse contexts.*

Keywords: verb phrase ellipsis, meaning in context, syntax-discourse interface, discourse comprehension, experimental pragmatics

* This material is based on work supported by the National Science Foundation under Grant No. BCS-1827404. We would like to thank the reviewers and editors of *Language* as well as Chris Kennedy, Jason Merchant, and the Language Processing Lab at the University of Chicago for their many helpful comments. All remaining errors are due to the authors.

1. INTRODUCTION. It is well established that language comprehension involves computing both the linguistic representations specific to an utterance and the larger discourse model in which the utterance is embedded. Nevertheless, we are far from a complete account of how structured linguistic representations are integrated into the broader mental model of the discourse, or how the discourse model is recruited during the interpretation of linguistic expressions. The current study concerns the empirical domain of elliptical expressions, in particular, verb phrase ellipsis (VPE), as shown by the example in 1.

(1) Don't!

Informally speaking, the elliptical expression in 1 is inherently 'incomplete', since the verb phrase is missing. The basic puzzle is how meaning is derived from an incomplete form. Such 'incomplete' expressions confront us with a fundamental question about the intricate relationship between form, meaning, and context. In the case of VPE, two competing observations obscure the nature of the interpretation strategy. The first observation is that VPE is often interpreted using the content of an antecedent verb phrase that was uttered in the recent linguistic context. At the same time, another well-known observation is that VPE can be acceptable and interpretable without a linguistic antecedent in a sufficiently informative context. We will review empirical findings regarding these observations below. These observations have given rise to heated debates about the extent to which the linguistic antecedent and broader context interact in determining the possible interpretations of VPE sites, and in particular, whether the linguistic antecedent, when present, is uniformly privileged over the broader discourse context as a locus of resolution.

In this paper, we present novel empirical evidence that constrains the range of possible interpretive mechanisms for VPE. Our experimental results indicate that salient information from the broader discourse context can influence the distribution of interpretations that arise for VPE sites even in the context of a viable linguistic antecedent, suggesting that the interpretive mechanism does not deterministically prefer resolution under exclusive identity with a local antecedent. At the same time, however, there is not a straightforward mapping between the propositions that are salient in the discourse at large and the interpretations that arise for ellipsis sites. Our findings suggest a 'hybrid' account of VPE interpretation that is largely guided by the form of the linguistic antecedent, but also takes into account the listener's assessment of the speaker's likely intended meaning in context.

2. BACKGROUND. A key debate in the literature on VPE discusses whether the relation linking ellipsis sites and their antecedents is predicated along the lines of syntactic or semantic identity.[1] This debate turns on one crucial type of evidence, namely whether the absence of a syntactically parallel linguistic antecedent leads to ungrammaticality for VPE sentences. In many cases, it does. For instance, in 2, the voice mismatch between the antecedent and the ellipsis site leads to an unacceptable sentence, suggesting a crucial constraint of syntactic identity.

(2) *This information was released, but Gorbachev didn't.

(Arregui et al. 2006, adapted from Hardt 1993)

Relatedly, the same island constraints that hold for fully realized sentences remain in effect when the position of the relevant trace is elided, as shown in 3. This has been taken as evidence that ellipsis sites contain unpronounced syntactic structure (Haïk 1987).

(3) *Abby wants to hire someone who speaks a Balkan language, but I don't remember which Balkan language Ben does. [want to hire someone who speaks t]

(Merchant 2004)

On the other hand, superficially similar sentences containing antecedent-ellipsis mismatches are marked as acceptable in the literature, indicating that a syntactically parallel antecedent is not always necessary for VPE to be acceptable. In 4, the antecedent is passive, while a fully realized interpretation of the ellipsis site after *did* is necessarily active.

(4) This problem was to have been looked into, but obviously nobody did.

(Kehler 2002)

In addition to voice mismatch, other kinds of antecedent-ellipsis mismatch are also possible. In 5, the antecedent VP contains the word *any*, which would be ungrammatical if the antecedent VP were adopted wholesale in the ellipsis site following *might*, yet the example is acceptable.

(5) We haven't decided to blacklist any firms. But there's a chance we might.

(Hardt 1993)

In split-antecedent situations, such as 6, the ellipsis site appears to receive an interpretation that is supported by two different antecedents; that is, the ellipsis site is interpreted as meaning that Wendy cannot sail around the world and that Bruce cannot climb Kilimanjaro.

(6) Wendy is eager to sail around the world and Bruce is eager to climb Kilimanjaro, but neither of them can because money is too tight.
(Webber 1978)

Proponents of syntactic identity between ellipsis sites and their linguistic antecedents point to examples like 2 to support the conclusion that ellipsis sites have internal syntactic structure, and that there is a requirement for this structure to be isomorphic to a linguistic antecedent for the sentence to be grammatical (Wasow 1972, Sag 1976, Williams 1977). The challenge for syntactic identity accounts is that they undergenerate; a strict interpretation of 'syntactic identity' would rule out acceptable mismatches like the voice alternation in 4 or the polarity alternation in 5.

One approach to reconciling a syntactic identity account with acceptable mismatch data is to separate the notions of grammaticality and acceptability for formal mismatches. According to the VP recycling hypothesis, presented by Arregui and colleagues (2006), the grammatical relation licensing VPE is syntactic identity with a linguistic antecedent, but VPE used in the absence of an isomorphic antecedent can be interpreted if the parser can recycle prior linguistic material to reconstruct a suitable linguistic antecedent. Their experimental results suggest that the level of acceptability of mismatch examples, which are ungrammatical on their model, is a function of the number and complexity of the syntactic operations required to reconstruct the VP antecedent. In a similar vein, Fox (2000), van Craenenbroeck (2012), and Thoms (2015) allow for accommodation of a syntactically isomorphic antecedent when VPE or a VPE-like construction is used in the absence of an isomorphic antecedent.

Another approach is to posit more articulated syntactic analyses of linguistic antecedents such that syntactic identity can be satisfied even in cases of apparent nonidentity. For instance, Merchant (2013) accounts for the acceptability of voice mismatches in VPE by positing a Voice head that is external to the phrase being elided, vP, meaning that VPs differing minimally in voice can be analyzed as syntactically identical. This contrasts with sluicing, where the phrase being elided is TP and necessarily contains the Voice head. Similarly, Kim and colleagues (2011) use a late merger operation to give a syntactic analysis of active and passive verb phrases according to which they are syntactically identical at some point in the derivation. Under this analysis, antecedent-ellipsis mismatch sentences are grammatical, but independently motivated

parsing heuristics invoked during the search for an antecedent, such as a preference for agents appearing in subject position, degrade their acceptability.

By contrast, proponents of a semantic identity account take examples like 4-6 as evidence that syntactic identity between an ellipsis site and its antecedent is not relevant to the acceptability of VP ellipsis. These accounts (e.g. Webber 1978, Dalrymple et al. 1991, Hardt 1993) treat VP ellipsis sites as phonologically null proforms that can receive their interpretation from a semantic antecedent, such as a property that has been linguistically evoked.

The general challenge for semantic identity accounts is that, on a simple interpretation of the model, they overgenerate with respect to the attested ellipsis data. For instance, it is not immediately apparent why 2 should be unacceptable, given that active and passive versions of the same sentence are generally analyzed as being semantically identical. Hardt (1993) proposes that the availability of ellipsis is restricted by the heuristics invoked to locate the semantic antecedent, such as a constraint favoring coreference between the subject of the antecedent and the subject of the ellipsis. In a similar vein, Kehler (2000, 2002) and Kertz (2008, 2013) argue that the range of possible VPE sentences is constrained by independent syntactic parallelism requirements that are necessary to establish the appropriate discourse coherence relation (Kehler) or that are motivated by information structural considerations linking the ellipsis site and the antecedent (Kertz).

While the discussion above focuses on whether there needs to be a formal identity relation between the ellipsis site and its linguistic antecedent, another type of empirical observation that can inform the mechanisms underlying the interpretation of VPE comes in the form of ellipsis used in the apparent absence of a viable linguistic antecedent. Antecedentless ellipsis has received less attention in the literature than cases like antecedent-ellipsis voice mismatch, in part because felicitous examples of this sort whose meanings are not idiomatically memorized are relatively rare. Hankamer and Sag (1976) point to a number of degraded examples of antecedentless ellipsis, such as 7, to argue that VPE sites cannot receive their interpretation from the nonlinguistic context.

(7) [Sag produces a cleaver and prepares to hack off his left hand.]
Hankamer: #Don’t be alarmed, ladies and gentlemen, we’ve rehearsed this act several times, and he never actually does.
(Hankamer & Sag 1976)

Subsequent work has indicated that antecedentless VPE does in fact occur productively, as in 8 and 9, even though it is much rarer than VPE used in the presence of a viable VP antecedent (Schachter 1977, Jacobson 2003, Merchant 2004, Miller & Pullum 2014).

(8) [Harry, alone in a corridor, discovers a classmate in an enchanted paralysis on the floor. Just then, the evil groundskeeper chances upon him, and, assuming Harry has laid the spell, runs to fetch a teacher. In a moment, he returns with the teacher, who shakes her head and turns away. Harry, aghast at being suspected of the vile deed, calls after her:]
I swear I didn't!
(Merchant 2004, from 'Harry Potter and the Chamber of Secrets', 2002)

(9) The aisles at the Lakewood Wal-Mart are surprisingly packed at 11 p.m. 'Can we? Can we?' Vanessa tugs at her mother, pointing to a rack of 'Lady and the Tramp' DVDs. Diaz shrugs. OK.
(Miller & Pullum 2014, from the Corpus of Contemporary American English)

If antecedentless VPE can attain the same level of acceptability as canonical VPE, then the question arises of how to account for its resolution with interpretive mechanisms that typically make reference to a linguistic antecedent. This is relatively straightforward for semantic accounts of interpretation; as long as the nonlinguistic context can provide the same type of semantic antecedent as an overt verb phrase, then antecedentless VPE should be able to pick out this antecedent in the same way canonical VPE is resolved under identity with a linguistic antecedent. This is the tack taken by Hardt (1993), who provides several examples of antecedentless VPE and suggests that they are resolved with respect to a salient property that is available in the nonlinguistic context. However, it is largely left unspecified what features of the nonlinguistic context can raise a property to sufficient salience to act as an antecedent.

Miller and Pullum (2014) provide a more detailed licensing account of both canonical and antecedentless VPE that relies on the availability of salient contrasts in the discourse at large. In the case of 'auxiliary-choice' VPE, for example, VPE is felicitous when the contrast between a particular proposition p and its negation is highly salient in the discourse, and the VPE utterance serves to pick out one of the alternatives as being the case. In canonical cases of ellipsis, this holds because an overt antecedent VP is used to assert, question, or otherwise discuss the possibility of p. They propose that as a rule, a nonlinguistic context on its own cannot assert or question the truth of a proposition such that it attains sufficient salience to license

antecedentless VPE. However, if an antecedentless VPE utterance specifically addresses the permissibility or advisability of a particular course of action, listeners may accommodate that a contrast between p and its negation was sufficiently salient in the discourse to license the utterance.

Antecedentless VPE on the surface appears to be more problematic for syntactic interpretation accounts. In their strictest form, these accounts require the presence of an antecedent verb phrase for a VPE site to be interpretable, and at best make no predictions about how VPE in the absence of such an antecedent would receive an interpretation. Merchant (2004) suggests a mechanism that allows for antecedentless ellipsis while maintaining an identity approach for canonical ellipsis. He proposes while that canonical ellipsis is interpreted under identity with a linguistic antecedent, hearers who encounter what appears to be an antecedentless VPE site infer that the elided material is actually a deep verbal anaphor such as *do it*. Listeners then resolve the anaphor using information from the discourse context at large, a process which is known to occur relatively freely for verbal anaphors of this kind (Hankamer & Sag 1976, Sag 1984, Merchant 2004).

To summarize, the hypothesis space of possible interpretive mechanisms for VPE sites can be viewed as a continuum. At one end of the continuum, syntactic identity is strictly privileged in the grammar. Any acceptable antecedent-ellipsis mismatch would be considered as the outcome of an accommodation mechanism, which is a secondary repair process triggered by some kind of error signal when no identical syntactic antecedent is present. The accommodation process in principle would apply either when the linguistic antecedent exhibits a mismatch with features at the ellipsis site, or when there is no linguistic antecedent at all. For convenience, we will call this kind of approach an IDENTITY STRATEGY.

At the other end of the continuum, what is at stake during interpretation is the availability of a semantic property that is salient in the discourse. Various constraints, whether syntactic, semantic, or discourse pragmatic, may play a role in determining what discourse property becomes the most salient and accessible to a listener, but once there is such a salient property available, it can license the ellipsis site and be targeted in interpretation. Under this kind of approach, an identical syntactic antecedent is not inherently more privileged than other sources of information. Again for convenience, we will call this a DISCOURSE-DRIVEN STRATEGY.

In principle, there is a relatively large hypothesis space on this continuum, depending on the specific parameters of different proposals. The current study uses a novel experimental paradigm to systematically investigate an empirical domain that has previously received little attention – discourse contexts in which the linguistic antecedent and the broader discourse context make different information available for the interpretation of an ellipsis site. The goal of the study is to use listeners' interpretation patterns in these contexts to derive new constraints on the hypothesis space for the VPE interpretive mechanism; that is, the experimental results will help narrow down analytical possibilities for VPE interpretation.

In Experiment 1, we look at the interpretation of VPE utterances in complex discourse contexts, where the linguistic antecedent and the discourse context at large make different information available for the interpretation of the ellipsis site. In particular, we examine whether, when the ellipsis site's linguistic antecedent is held constant, manipulations in contextual information available outside the antecedent VP can affect the distribution of interpretations available for the ellipsis site. This addresses the question of whether VPE interpretation always considers information from outside the antecedent VP, or whether such information is ignored in cases where the ellipsis site can be resolved under identity with a linguistic antecedent. Second, in Experiment 2, we quantitatively assess which propositions are most salient in the Experiment 1 contexts and compare the results to the interpretations available for VPE.

3. EXPERIMENT 1: VPE INTERPRETATION IN CONTEXT. The goal of Experiment 1 was to systematically assess the interpretation of VPE in complex discourse contexts, where the local antecedent VP and the broader discourse context contribute potentially conflicting information for inclusion in the interpretation. Previous work on identity-driven accounts has generally paid little attention to information available in the broader discourse, including the nonlinguistic context. On the other hand, discourse-driven accounts of VPE interpretation have left the interaction of the linguistic and broader discourse contexts in determining the ellipsis site's interpretation largely unspecified. Experiment 1 introduces a missing data point to the discussion, namely cases where VPE is used in the presence of both a viable linguistic antecedent as well as a broader discourse context that makes potentially relevant information salient.

To briefly preview, the critical measure in Experiment 1 is whether participants interpret a VP ellipsis site as containing a numeral. The experimental stimuli vary the level of support for number information in both the linguistic antecedent and the broader discourse context. Of interest is whether, in the presence of a linguistic antecedent, manipulations in the broader discourse context affect the distribution of interpretations. If these broader manipulations are ignored in such contexts, it suggests that the VPE interpretation mechanism deterministically prefers resolution under identity when a VP antecedent is available. If an effect of broader discourse manipulations is seen, it suggests a more complex interpretation mechanism that considers information from multiple sources in the discourse.

At the grocery store

Son: **I want to buy candy bars!**

Father: **We can't.**

Based on the scenario above, do you think it is more likely that **the father** meant:

1. We can't buy five candy bars, but maybe we could buy fewer.
2. We can't buy any candy bars.

Use number keys or click choice to answer.

FIGURE 1. Sample Experiment 1 trial.

3.1. DESIGN. Each trial in the experiment was composed of a comic strip context, an (optional) antecedent utterance, a VPE reply utterance, and an interpretation prompt. The comic strip context was displayed at the top of the screen under a scenario title, while the utterances were displayed as text below the comic strip context with an italicized label indicating the character who was speaking. The interpretation prompt and answer choices were shown below this, and participants could select a choice by clicking on it or by pressing a number key on their keyboard. A sample experimental trial is shown in Figure 1.

The experiment featured a 3×3 (comic strip context × linguistic antecedent) design. The three levels of comic strip context were UNAVAILABLE, AVAILABLE, and SALIENT. The strips differed in the amount of number information they provided about a referent mentioned in the antecedent utterance. The number information could be absent (unavailable); represented, but not made salient, by a character interacting with the referents as a group (available); or made salient by a character interacting with referents one by one and counting them (salient). Sample comic strip contexts are shown in Figure 2.

The levels for the linguistic antecedent were EXOPHORIC, UNMODIFIED, and MODIFIED. In the exophoric conditions, there was no antecedent utterance, so only the VPE reply was shown under the comic strip context. The unmodified antecedents were utterances with no numeral in the VP, attributed to the character interacting with the referents. For example, in the scenario shown in Figure 1, the antecedent *I want to buy candy bars!* is attributed to the son. The modified antecedent is the same as the unmodified antecedent, except that the numeral relevant to the scenario is added to the VP, such as *I want to buy five candy bars!*

The VPE reply utterance is attributed to the other character. It is an elliptical response that constitutes a negative comment on the permissibility or advisability of the suggested course of action, predicted by Miller and Pullum (2014) to be acceptable even in exophoric contexts as long as the discourse support is sufficiently strong. The reply utterance in the example scenario, attributed to the father, is *We can't*.

Below the comic strip context and utterances, subjects were prompted to select whether they thought one of two interpretations, POLAR or SCALAR, was more likely as the intended interpretation of the VPE reply. The polar interpretation corresponds to interpretation of the ellipsis site with no numeral, for example, *We can't buy [any] candy bars*. The scalar

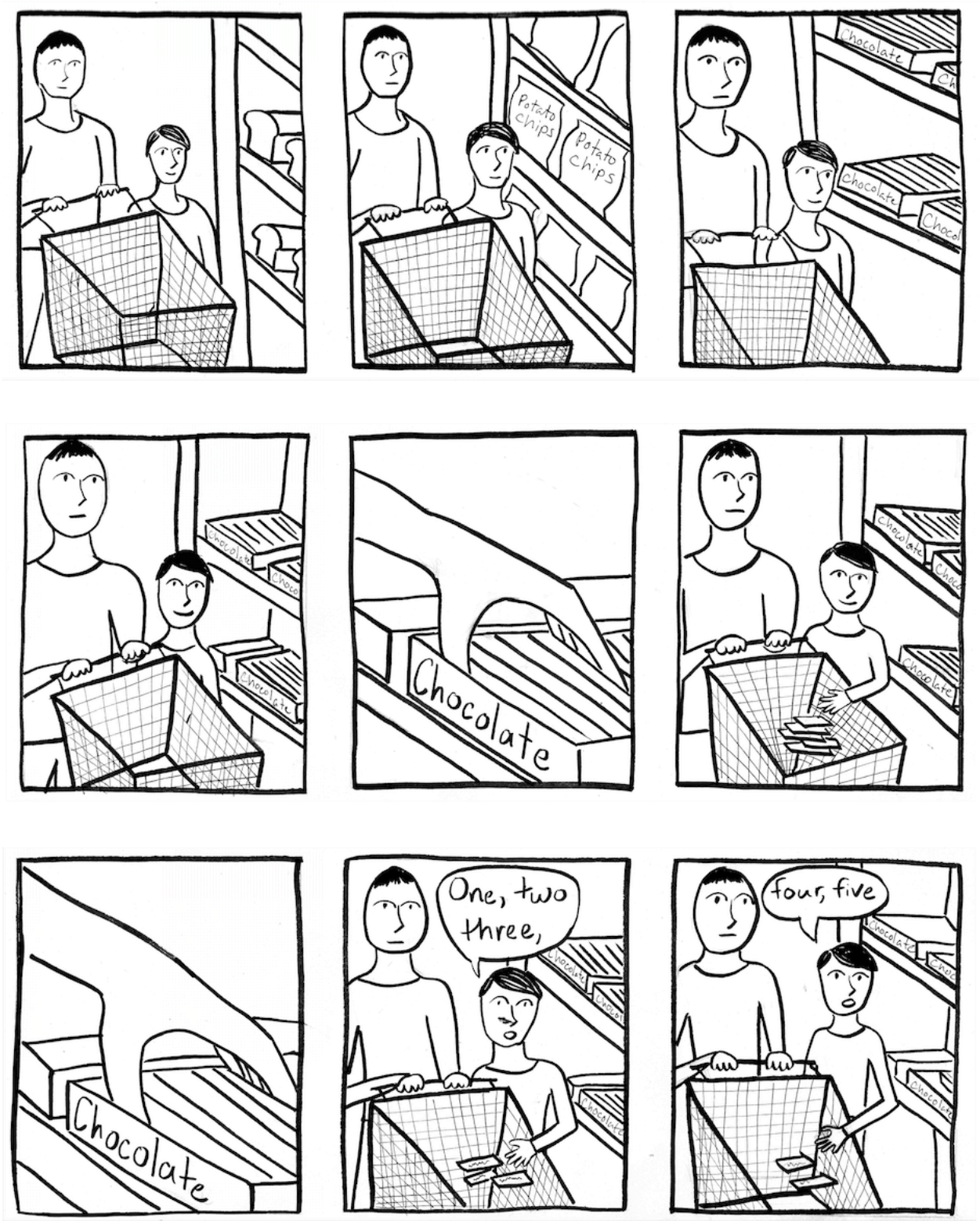


FIGURE 2. Top to bottom: unavailable, available, and salient comic strip contexts for a sample Experiment 1 item.

interpretation represents interpretation of the ellipsis site with a numeral, for example, *We can't buy five candy bars, but maybe we could buy fewer*. The order of the interpretations was randomized by trial.

Thus, there were 9 experimental conditions consisting of a comic strip context and an antecedent-reply discourse, followed by the polar-scalar interpretation prompt. The sample scenario in each condition is described in Table 1.

	Exophoric	Unmodified antecedent	Modified antecedent
Unavailable context	**Context:** [Son stands near candy bars.] **Father:** We can't.	**Context:** [Son stands near candy bars.] **Son:** I want to buy candy bars! **Father:** We can't.	**Context:** [Son stands near candy bars.] **Son:** I want to buy five candy bars! **Father:** We can't.
Available context	**Context:** [Son takes five candy bars at once.] **Father:** We can't.	**Context:** [Son takes five candy bars at once.] **Son:** I want to buy candy bars! **Father:** We can't.	**Context:** [Son takes five candy bars at once.] **Son:** I want to buy five candy bars! **Father:** We can't.
Salient context	**Context:** [Son counts out five candy bars.] **Father:** We can't.	**Context:** [Son counts out five candy bars.] **Son:** I want to buy candy bars! **Father:** We can't.	**Context:** [Son counts out five candy bars.] **Son:** I want to buy five candy bars! **Father:** We can't.
Prompt: Based on the scenario above, do you think it is more likely that **the father** meant: **Polar:** We can't buy any candy bars. **Scalar:** We can't buy five candy bars, but maybe we could buy fewer.			

TABLE 1. Conditions and interpretation prompts for a sample item of Experiment 1.

3.2. PROCEDURE. Participants were recruited using Amazon Mechanical Turk, while the experiment itself was executed in participants' browsers using the Ibex Farm platform (Drummond 2020). Participants provided informed consent and completed a demographic questionnaire and two unrelated practice items before beginning the experiment itself.

Participants were told to look at the comic strip to familiarize themselves with the scenario, then read both utterances silently. They were told they could respond to the prompt by clicking on the choice or pressing the corresponding number key, and either of these actions caused the next item to display immediately. The experiment took about 15 minutes to complete and participants were paid USD 1 for their participation.

3.3. STIMULI. Six critical scenarios were constructed in nine conditions on the model of the example scenario described above. The critical information manipulated in the comic strip

context and the antecedent utterance was always about a particular number between three and five that varied by item. The VPE replies for each item are similar to the example scenario above, with minor adjustments, as needed, in the pronouns and in the modal auxiliary (e.g. *shouldn't*, *can't*).

An experimental session consisted of six critical trials spanning the six critical scenarios, and ten filler items that involved VPE interpretation, but without the numeral manipulation. For each participant, the six critical trials were drawn from different conditions of the 3×3 design, such that each participant gave one response each in six of the critical conditions, and was not exposed to the other three conditions. When all data was considered together, the number of observations of each scenario and each condition was roughly evenly distributed, with between 109 and 134 observations retained in each condition.

3.4. PARTICIPANTS. 198 participants (85 female, mean age 34.4) over the age of 18 participated in the experiment. 13 subjects were excluded from analysis because they failed to correctly answer at least two out of three filler items that were designed to measure participant attention. 4 additional subjects were excluded from analysis because they failed to clearly identify themselves as native English speakers in the demographic survey.

3.5. RESULTS. The proportions of polar responses by condition are shown in Figure 3. Note that for each condition, the proportion of scalar responses is 1 minus the proportion of polar responses.

A visual examination of the plot suggests that both the antecedent manipulation and the comic strip context manipulation had an effect on the proportion of polar versus scalar responses. In all three antecedent conditions, the proportion of times the polar interpretation was selected appears to decrease as a function of the number information available in the comic strip context. However, this effect is mediated by the antecedent. In the exophoric conditions, a preference for the polar interpretation in the unavailable context gives way to a preference for the scalar interpretation in the available and salient contexts. In the unmodified conditions, the polar interpretation is always preferred, although its advantage erodes with increasing number information in the comic strip context, and the proportions of polar and scalar responses are close with a salient context. With a modified antecedent, the scalar interpretation is preferred even

with an unavailable context, and its advantage increases as a function of number information in the comic strip context.

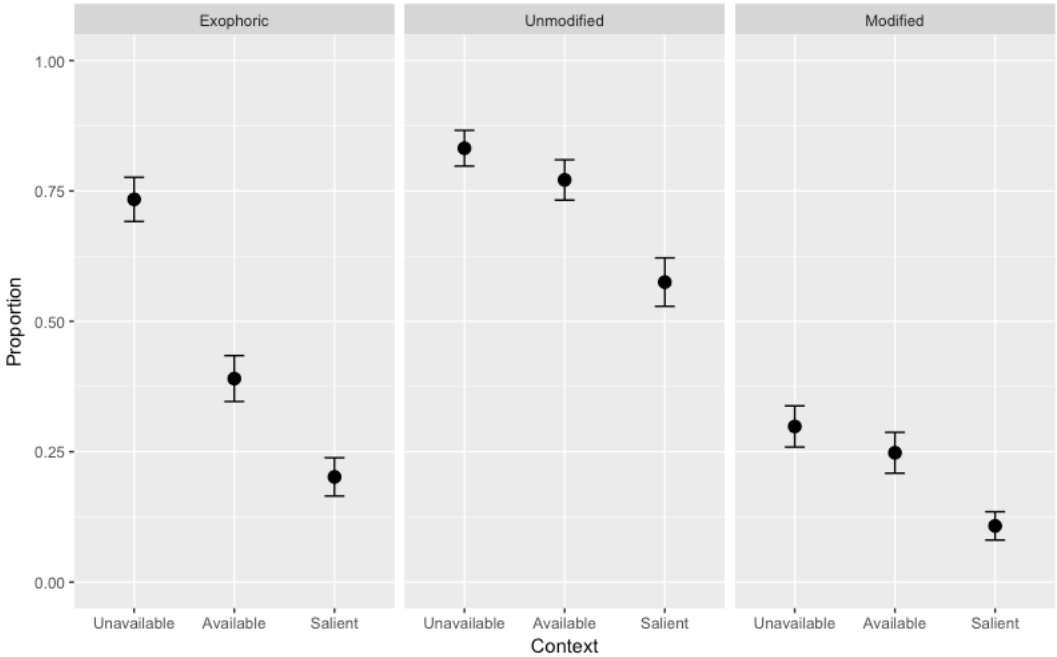


FIGURE 3. Experiment 1 results: proportions of polar responses. The error bars represent by-participant standard errors.[a]

3.6. ANALYSIS. The response data were coded with polar responses as 1 and scalar responses as 0. Two logistic mixed effects regression models were constructed using the lme4 package in R (Bates et al. 2015). The first model predicted the response data with fixed effects for comic strip context, antecedent, and their interaction. This model featured the maximal convergent random effects structure: random intercepts for participants, and a random intercept for item and random slopes for item in context and antecedent.[2] The second model had the same specification with the exception of the interaction of comic strip context and antecedent. Model comparison showed that the model with the interaction was a significantly better fit for the data ($p<.01$). On the basis of the significant interaction, further analysis of the effect of comic strip context on the proportion of polar responses was conducted on a by-antecedent basis.

[a] In both experiments, the by-participant standard errors are equivalent to the standard errors calculated over the raw pooled data, since each participant saw each condition at most once.

For each antecedent condition, we performed two types of analyses. First, we constructed logistic mixed effects regression models for the data from each antecedent condition with a main effect of context, random intercepts for participant, and random intercepts for item and random slopes for item in context.[3] We examined whether there was a gradual change in the polar response proportions when contextual support for number information increased from unavailable to available and then to salient. For this analysis, contextual support was coded using backward difference coding, with the unavailable context coded as the baseline level. With this coding scheme, the proportion of polar responses at each context level was compared to the proportion of the previous level.

In the exophoric conditions, the proportion of polar responses in the available condition was significantly different from the proportion in the unavailable condition, and the proportion in the salient condition was significantly different from that in the available condition (p's$<.01$). In the unmodified antecedent conditions, the proportion of polar responses was not significantly different in the unavailable and available conditions ($p>.2$), but the proportions in the salient and available conditions were significantly different ($p<.001$). Similarly, in the modified conditions, the proportions were not significantly different for unavailable and available strips ($p>.3$), but were significantly different between the available and salient comic strip contexts ($p<.05$).

We also conducted a separate analysis to determine whether in each condition participants significantly preferred one interpretation over the other. A logistic mixed model with random intercepts for item was constructed for the data from each antecedent/context condition.[4] The proportion of polar responses was significantly different from 0.5, indicating a significant preference for one of the two interpretations, for every condition besides unmodified/salient ($p>.1$) at the following significance levels: exophoric/available, $p<.05$; exophoric/unavailable, $p<.01$; all other conditions, $p<.001$.

3.7. DISCUSSION. Experiment 1 tested how VPE is interpreted in a rich discourse context, where material from both the linguistic antecedent and the discourse at large are made available (to various degrees) during interpretation. It is clear from the previous literature that for the exophoric conditions, even a strict identity account must allow for some ‘backup’ mechanism for interpretation with respect to the broader discourse. We therefore expected to see a strong effect of context in the antecedentless conditions. What is more interesting are the conditions that have

a grammatically licit antecedent. It is these conditions that offer a testing ground for the potential interaction between the linguistic antecedent and the broad discourse context.

Turning to the results, in the exophoric conditions, VPE interpretation was largely a function of the availability of number information in the comic strip context. When we tested whether there was a significant preference for the polar or scalar interpretation in each condition, we found that with an unavailable context, the polar interpretation was significantly preferred. As number information became more available in the context, though, the scalar interpretation gained strength and was preferred with both the available and salient contexts. This finding is further corroborated by the analysis done for each antecedent type, which directly tested whether switching from the unavailable to the available context or from the available to the salient context significantly altered the proportion of polar responses. This analysis showed a significant difference in the polar response proportions for the unavailable and available conditions and for the available and salient conditions, suggesting that the strength of the polar interpretation decreased in response to increasing contextual availability of number information. This pattern of results within the exophoric conditions is unsurprising; since there was no linguistic antecedent in these conditions, subjects had to use the broader context to resolve the ellipsis, and the interpretation closely tracked the information that was available in the comic strip context. This is in line with previous work on antecedentless ellipsis (Merchant 2004, Miller & Pullum 2014), which suggests that in the absence of a linguistic antecedent, VPE must be interpreted with respect to the information generally available in the discourse context.

The conditions with antecedents (i.e. the unmodified and modified conditions) likewise show effects of context. In the unavailable context, where there was little numeral information in the broader context, the antecedent-faithful interpretations were clearly preferred. In particular, the polar interpretation was preferred when the antecedent was unmodified, and the scalar interpretation was preferred when the antecedent was modified. Based on the analysis that examined the interpretation differences from one context to the next, switching from the unavailable to the available context did not significantly shift the interpretation preferences. Changing to the salient context, however, had a clear effect. In the unmodified-antecedent condition, the antecedent-nonidentical interpretation, the scalar interpretation, received a significant boost and was treated as approximately equally available as the antecedent-identical interpretation. In the modified-antecedent condition with a salient context, the number

information from both the comic strip context and the linguistic antecedent led to an even stronger preference for the scalar interpretation compared to the other contexts.

The critical findings from the current results have been reproduced in another study using a 1-7 Likert scale for rating the two VPE interpretations separately (Geiger & Xiang 2017, Luce et al. 2018). Altogether, the results of Experiment 1 are problematic for a VPE resolution mechanism at the far 'identity' side of the continuum. According to the identity strategy, strictly formulated, information from the discourse outside the linguistic antecedent should not play a role in ellipsis interpretation when the linguistic antecedent can readily be used to interpret the ellipsis site. However, in the unmodified and modified conditions, the proportion of polar versus scalar responses actually did vary in response to changes in the comic strip context.

Thus, Experiment 1 has provided a first constraint on the space of plausible hypotheses for the mechanism of interpreting VPE in context, namely that ellipsis sites situated in complex discourse contexts are not interpreted under exclusive identity with a local VP antecedent. The results are certainly compatible with an account that makes use of discourse information to resolve VPE interpretation, but a number of possibilities remain open. In one possibility, VPE interpretation would be solely determined by the salience or accessibility of different properties made available by the context. This would predict that a listener's preference for a particular VPE interpretation should be in proportion to the strength of the corresponding proposition in context. On the other hand, it is also possible that broad discourse context constrains the VPE interpretation through modifying possible candidate interpretations, but it cannot completely undo the effect of the linguistic antecedent. To tease these possibilities apart, we need to first have a way to quantitatively assess the prior availability of different propositions in the context before the VPE utterance was made. We carry out this task in Experiment 2.

4. EXPERIMENT 2: EMPIRICALLY ASSESSING THE DISCOURSE-SALIENT PROPOSITIONS.

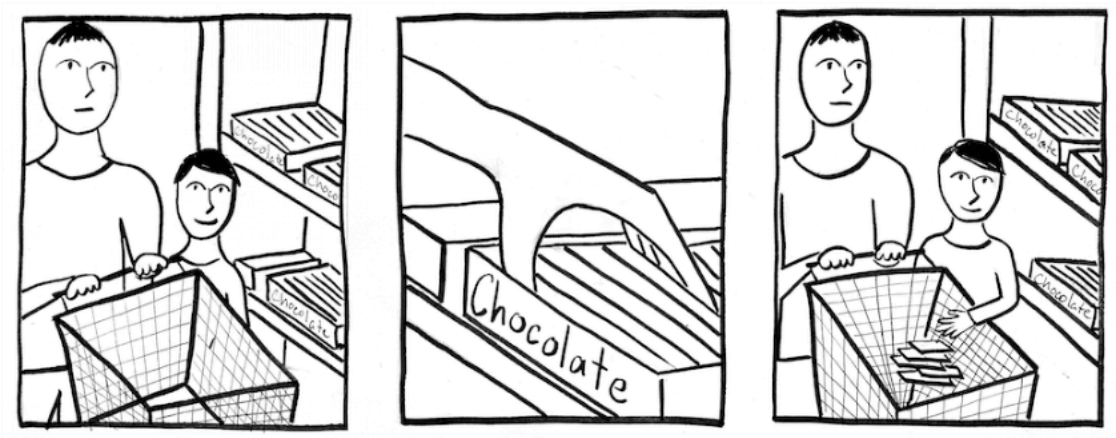
Experiment 2 assessed the general salience of a range of propositions in the Experiment 1 discourse contexts prior to the utterance of the VPE reply. The main goal of the experiment was to measure the accessibility of the propositions corresponding to the polar and scalar VPE interpretations and determine whether they map straightforwardly onto the availability of these interpretations under ellipsis.

4.1. DESIGN, PROCEDURE, AND STIMULI. Experiment 2, including the procedures for informed consent, recruitment, and payment, was identical to Experiment 1 with two critical modifications. First, the VPE reply was deleted, so participants were evaluating a discourse consisting of a comic strip context and one optional utterance (the Experiment 1 antecedent). Second, the interpretation prompt was changed to ask which of three possible 'mental states' were most likely for the character that utters the antecedent. The mental states provided were INDIFFERENT (*The son wants to buy candy bars, but doesn't care how many*), SPECIFIC (*The son wants to buy a specific number of candy bars*), and ZERO (*The son does not want to buy any candy bars*). The filler items were also adjusted to remove VPE utterances and pose a comparable question. A sample trial from Experiment 2 is shown in Figure 4.

The main goal was to assess what a listener would perceive as the most salient 'issue' in the discourse before the VPE utterance was uttered. Under a discourse-driven account for VPE interpretation, there is at least a partial mapping between what is at issue in the discourse context and the interpretation a VPE utterance receives. If a listener decides, after hearing the antecedent sentence, that what is at issue is buying candy bars, regardless of number (indifferent), this supports the conclusion that the VPE reply *We can't* means that no candy bars can be purchased – the polar interpretation. If the listener decides that buying a specific number of candy bars is at issue (specific), then *We can't* would be more likely to mean that that particular number of candy bars can't be purchased, which is the scalar interpretation. Thus, we treat the indifferent and specific mental states as proxies for the general discourse availability of the polar and scalar VPE interpretations, respectively.

The third choice, described above as the zero mental state, was included in Experiment 2 only to maintain the naturalness of the experimental task, as it might be the most natural choice in the exophoric/unavailable condition. However, we do not believe this interpretation of the discourse is relevant to the interpretation of the VPE reply in Experiment 1, as it is not compatible with the reply *We can't*; that is, *We can't* does not seem like a felicitous reply when what is at issue is buying zero candy bars. For the sake of parallelism with Experiment 1, the results and analysis below present the data with the zero responses excluded. There were a considerable number of zero responses in the exophoric/unavailable condition (28.7% of responses), while the number of zero responses in the remaining conditions was substantially lower (between 2.2% and 10.0% of responses by condition).[5]

At the grocery store



Son: **I want to buy five candy bars!**

Based on the scenario above, which of the following do you think is most likely?

1. The son wants to buy a specific number of candy bars.
2. The son does not want to buy any candy bars.
3. The son wants to buy candy bars, but doesn't care how many.

Use number keys or click choice to answer.

FIGURE 4. Sample Experiment 2 trial.

4.2. PARTICIPANTS. 227 participants (103 female, mean age 34.8) over the age of 18 participated in Experiment 2. 22 subjects were excluded from the analysis for failing to correctly answer at least two out of three attention trials, and a further 2 were excluded for not self-reporting as native English speakers.

4.3. RESULTS AND ANALYSIS. The proportions of indifferent responses, calculated after discarding the zero responses, are plotted in Figure 5. Note that the proportion of specific responses is thus 1 minus the proportion of indifferent responses. Since the primary goal of Experiment 2 is to assess if discourse salience alone can explain the results from Experiment 1, our analysis will focus on the comparison between the two experiments. In Figure 5, we overlay the proportions of polar responses from Experiment 1 on the proportions of indifferent responses from Experiment 2.

Before comparing the results from the two experiments, we first conducted an analysis on the data from Experiment 2 alone. As in Experiment 1, a logistic mixed model with a random intercept for item was separately constructed for the data from each antecedent/context condition to test whether there was a preference for one of the two mental states in each condition. The proportion of indifferent responses was significantly different from 0.5, indicating a preference for one of the two mental states, in the following conditions for the accompanying significance levels: unmodified/unavailable, $p<.05$; modified/unavailable, $p<.01$; exophoric/salient, unmodified/salient, modified/available, modified/salient, $p<.001$. The difference from 0.5 was not reliable in the unmodified/available condition ($p<.1$), the exophoric/unavailable condition ($p>.1$), or the exophoric/available condition ($p>.3$).

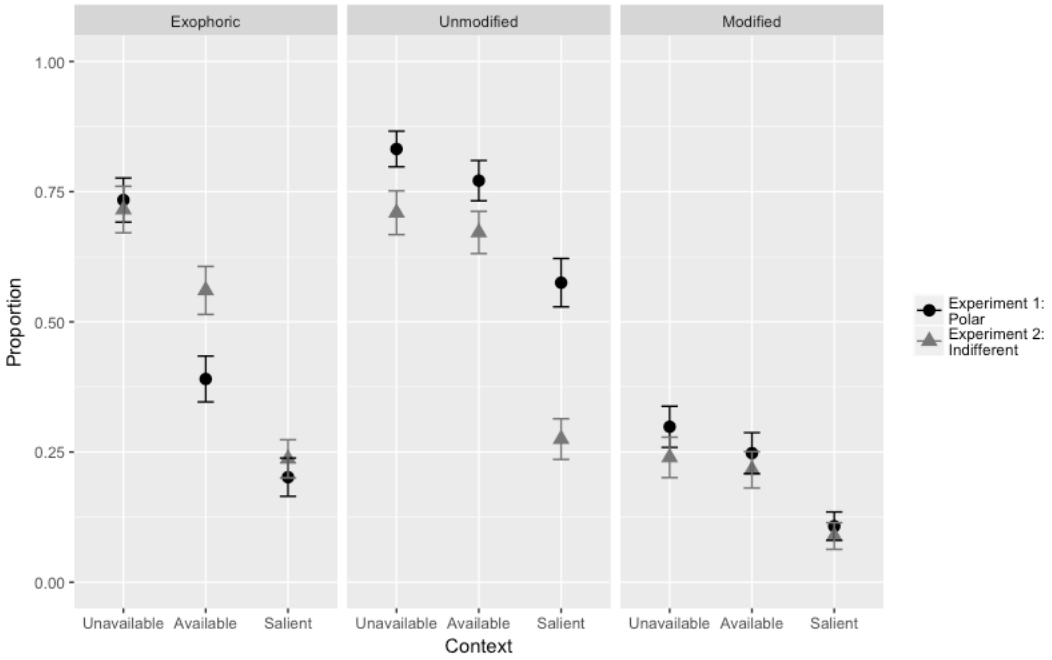


FIGURE 5. Experiment 1 and Experiment 2 results: proportions of polar and indifferent responses, respectively. The error bars represent standard errors of by-participant means.

Turning to the comparison between the two experiments, qualitatively speaking, there is a strong correspondence between the propositions available in the discourse at large and the interpretations available for VPE. For example, in both the exophoric and the modified conditions, the preference for a given VPE interpretation (i.e. polar versus scalar) closely tracks the salience of the corresponding discourse proposition (i.e. indifferent versus specific,

respectively). As shown in Figure 5, under these conditions, as the availability of the indifferent mental state decreased, the acceptability of the polar VPE interpretation also decreased. However, there is also one major misalignment between the two experiments, in the unmodified conditions: while the preference for the indifferent mental state sharply dropped from the unavailable and available contexts to the salient context, the availability of the corresponding polar VPE interpretation did not degrade to the same degree.

To quantitatively evaluate the difference between the two experiments, especially with respect to the unmodified/salient condition, we conducted Bayesian hierarchical analyses using the brms package from R (Bürkner 2017). Before the analysis, we implemented a user-defined contrast coding scheme for both the context and antecedent variables. For context, we constructed the contrast to compare the salient level to the mean of the unavailable and available levels. For antecedent, we set up the contrast to compare the unmodified level to the mean of the exophoric and modified levels. This coding scheme allowed us to directly test whether the change in experiment had a unique effect on ratings in the unmodified/salient condition compared to the other conditions of the experiment. The experiment variable was treatment coded, with Experiment 1 as the baseline level.[6]

We first analyzed the subset of the data from only the unmodified-antecedent conditions (6 conditions total, middle panel in Figure 5). The statistical model included fixed effects of context, experiment, and their interaction, and the full random effect structure including both by-item and by-participant intercepts and slopes.[7] There was strong evidence for an interaction between context and experiment (estimate = -0.88, SE = 0.46, 95% CrI [-1.95, -0.16]), suggesting that the difference between the unmodified/salient condition and the other two unmodified conditions was larger in Experiment 2 than in Experiment 1.

In a second model, we examined the subset of the data from the salient-context conditions (6 conditions total). The model setup was parallel to the first model with the antecedent variable in place of context, with fixed effects of antecedent, experiment, and their interaction, and the maximal random effect structure.[8] Here, there was also strong evidence for an interaction between antecedent and experiment (estimate = -0.95, SE = 0.42, 95% CrI [-1.90, -0.28]), driven by the fact that the difference between the unmodified/salient condition and the other two salient conditions is smaller in Experiment 2 than in Experiment 1.

Ultimately, both interactions observed above reflect the fact that there is a uniquely large difference between Experiments 1 and 2 in the ratings for the unmodified/salient condition, with the polar interpretation in this condition of Experiment 1 rated much lower than the corresponding indifferent mind state in Experiment 2. This effect was confirmed by a separate analysis comparing the ratings for the two experiments in only the unmodified/salient condition, with a fixed effect for experiment and the maximal by-item random effect structure.[9] Again, there was strong evidence for a difference between the two experiments (estimate = -1.24, SE = 0.42, 95% CrI [-2.03, -0.32]. Finally, we also checked whether the three-way interaction of context, antecedent, and experiment would hold by running a model on the complete data set.[10] There was some evidence for an effect of a three-way interaction (estimate = 0.16, SE = 0.1, 95% CrI [-0.37, 0.01]).

4.4. DISCUSSION. The goal of Experiment 2 was to estimate the most salient 'issues', or propositions supported by the broad discourse context prior to the Experiment 1 VPE utterance. The simplest possible discourse-driven model of VPE interpretation would be one that predicts a straightforward one-to-one mapping between a proposition's prior discourse salience and its availability under ellipsis. This prediction was borne out in the majority of the experimental conditions, where there was a close correspondence between the proportion of polar responses in Experiment 1 and the proportion of indifferent responses in Experiment 2. However, this simple discourse-driven analysis is nonetheless ruled out by a comparison of the results of Experiments 1 and 2 in the unmodified/salient condition. The Bayesian analysis comparing the two experiments showed that in the unmodified/salient condition, the indifferent mental state was rated as much less available in Experiment 2 than the corresponding polar VPE interpretation in Experiment 1 (in other words, the number-modified specific mental state was rated as significantly more available in the discourse in Experiment 2 than the corresponding scalar interpretation for VPE in Experiment 1).

This result indicates that a VPE interpretive mechanism invoking only prior salience considerations is inadequate as a model of interpretation in context. According to this strategy, because the specific mind state is most accessible before the VPE utterance, the scalar interpretation should be the preferred interpretation of the VPE reply in Experiment 1, contra the actual observation in Experiment 1. The most obvious explanation is that there is a bias in favor

of interpretations that are faithful to the content of the linguistic antecedent, and that this bias inhibits consideration of even very salient antecedent-nonidentical information from the broader discourse context. As a result, the antecedent-identical polar interpretation was preferred in spite of the greater availability of a number-modified proposition in the discourse at large.

Thus, the results of Experiments 1 and 2, considered together, appear to suggest that neither a strict identity strategy nor a purely salience-driven discourse strategy is sufficient to capture the facts of VPE interpretation in context. We turn to the implications of these findings in the next section.

5. GENERAL DISCUSSION. The experiments presented above were designed to narrow down the set of plausible mechanisms for context-situated VPE interpretation. Experiment 1 tested whether interpretation proceeds exclusively under identity with a linguistic antecedent. Experiment 2 investigated the degree to which the available interpretations under ellipsis are predicted by the availability of different propositions in the prior discourse at large.

Experiment 1 showed that in the context of a viable linguistic antecedent, the interpretations that arise for an ellipsis site are nevertheless affected by manipulations in the information that is available in the discourse context beyond the local VP antecedent. With an unmodified antecedent, which did not contain a numeral, participants relatively rarely selected the antecedent-nonidentical scalar interpretation in the unavailable and available contexts, where number information is unavailable or only weakly available in the discourse context. However, in the salient context, where number information is strongly evoked in the broader context, participants became much more likely to select the scalar interpretation. Similarly, with a modified antecedent, while the antecedent-identical scalar interpretation was already preferred with an unavailable or an available context, the inclusion of strong number information in the broader context in the salient condition led to even more selection of the scalar interpretation.

The Bayesian analysis following Experiment 2 indicated that although there is a close relationship between the propositions that are highly salient in the general discourse prior to the use of VPE and the interpretations that arise for an ellipsis site, there is also considerable mismatch when the results from the two experiments were analyzed together. In particular, in the condition with an unmodified antecedent and a salient context, we found evidence for a

substantial mismatch in the availability of certain propositions in the discourse at large and their availability as VPE interpretations.

The experimental findings discussed above constrain the range of possible interpretive mechanisms for VPE. First, the observation that manipulations in the information available in the broader discourse context affects VPE interpretation in the presence of a viable linguistic antecedent rules out any account involving an antecedent selection mechanism that gives deterministic preference to resolution under identity with a linguistic antecedent. Second, the mismatches between the propositions available in the discourse at large and the interpretations that arise for VPE also rule out approaches under which the distribution of interpretations for a VPE site is straightforwardly a function of the prior availability of the relevant propositions in the discourse at large. This includes accounts where listeners deterministically resolve a VPE site as addressing the most salient proposition in the context, or are proportionally likely to interpret VPE as addressing a particular proposition according to its prior availability in the context.

The current results extend the findings from previous literature. It has long been acknowledged that a strict identity account of VPE licensing undergenerates in that it categorically rules out mismatch sentences that are canonically marked as acceptable. In response to this, a number of mechanisms have been proposed that allow for the accommodation or reconstruction of an alternative antecedent (Fox 2000, Arregui et al. 2006, Thoms 2015). Each of these accounts retains a grammatical requirement for identity between the antecedent and the ellipsis site. Under certain conditions, when there is a formal mismatch between the two, a covert alternative antecedent that satisfies the identity requirement can be accommodated. Our findings suggest that even these augmented identity accounts are inadequate to fully capture the facts of context-situated VPE interpretation. Each of these models requires the presence of a formal antecedent-ellipsis mismatch to trigger the accommodation operation. By contrast, our Experiment 1 data show an effect of the broader discourse context on VPE interpretation even in the presence of a formally compatible linguistic antecedent. In other words, where even the most permissive existing identity accounts predict interpretation under strict identity in Experiment 1, we actually observed an effect of discourse.

Our data also show that there is not a straightforward mapping between the propositions that are available in the discourse at the time of utterance of a VPE sentence and the interpretations that are available for the ellipsis site itself. This is in line with some previous

discourse accounts, which acknowledged that interpretation goes beyond merely accessing the most accessible information in the discourse. For instance, in Hardt 1993, although the accessibility of the discourse information is highly relevant, there are still certain heuristics that guide the ultimate selection as to which antecedent property should be recruited for VPE resolution. The current study, by systematically manipulating both the linguistic antecedent and the discourse context, provides quantitative evidence clearly showing the interaction between the narrow grammatical antecedent and the broader discourse context. We note that the interpretations supported by the antecedent and the broader context can be mutually reinforcing, as in the modified conditions, where the scalar interpretation is always preferred, but it is preferred to a significantly higher degree when number information is highly available in the comic strip context. Information from the two sources can also exhibit a mutually subtractive effect. For instance, in the unmodified conditions, the antecedent-supported polar interpretation is generally preferred, but the addition of salient number information in the broader context erodes this preference such that participants are roughly indifferent between the polar and scalar interpretations.

Altogether, our results provide some characterization of the features required of a successful account of VPE interpretation in context. The most apparent of these is a strong but partially defeasible preference for the interpretation(s) that would be supported were the antecedent VP to be adopted wholesale in the ellipsis site. This preference is ‘strong’ in the sense that participants seem to still prefer antecedent-supported interpretations even in contexts where the nonidentical interpretation is significantly more available in the discourse, such as the unmodified/salient condition of Experiment 1. It is ‘partially defeasible’ in that broader discourse information can erode participants’ likelihood of selecting the antecedent-supported interpretation, but not so much that it is significantly dispreferred to a competing, context-supported interpretation.

Consistent with the observation of the additive and subtractive interaction of the linguistic and broader discourse contexts, we propose that VPE interpretation proceeds primarily according to the linguistic antecedent, but that listeners use information from the broader discourse context to modulate their interpretation. It is not a controversial idea that listeners derive the meaning of an utterance by not only considering the form of the utterance, but also the communicative intent of the speaker in the discourse context. It has long been recognized that the

listener's reasoning about the contextualized communicative intent of the speaker is a critical component of naturalistic language comprehension (Grice 1975), and more recent work in probabilistic pragmatics has suggested that both speakers and hearers may perform sophisticated, iterative modeling of each other's production and interpretation strategies (Bergen & Goodman 2015, Frank & Goodman 2012, Franke 2009, Goodman & Stuhlmüller 2013, Jäger 2011, Lassiter & Goodman 2017). Nevertheless, it is worth considering why listeners consult the broader discourse even in contexts where the linguistic antecedent alone would be sufficient to guide interpretation.

One possibility is that VPE interpretation – or utterance interpretation generally – involves some degree of uncertainty, and consideration of broader discourse information is part of a mechanism that helps listeners cope with this uncertainty. Work on the noisy channel model of sentence comprehension suggests that a major component of a listener's task during interpretation is the management of uncertainty regarding the signal transmitted by the speaker and the communicative intent underlying it (Levy 2008, Gibson et al. 2013, Futrell & Levy 2017, Ryskin et al. 2018). While this work primarily deals with cases where part of the transmitted signal is obscured, it is possible that a core component of utterance interpretation is a general uncertainty about speakers' intended messages in context. Our experimental results invite an analysis where VPE sites have clearly supported interpretations associated with the antecedent VP, but listeners nonetheless check these interpretations against the discourse at large to determine whether they are likely meanings the speaker could have intended. In cases where the antecedent-supported interpretation matches the discourse well (e.g. the unmodified/unavailable and modified/salient conditions), listeners feel more certain that the antecedent-supported interpretation is correct, and there is little consideration of alternative conclusions. However, in situations where the broader discourse context suggests that something other than the antecedent-supported reading might be at issue (e.g. the unmodified/salient condition), listeners become less certain that the antecedent-supported meaning is the one intended by the speaker, and they consider alternative readings that incorporate salient information from the broader context. Critically, this consideration of discourse context information is not necessarily a secondary accommodation process triggered by the absence of a grammatically licit antecedent. The current results are most compatible with a model in which listeners regularly recruit discourse information to assess the communicative intent of the speaker.

We also want to point out that although it appears to be a general strategy for listeners to consult the discourse context when interpreting VPE, there are crucial differences between VPE and other verbal anaphors in terms of the relative contribution of antecedent versus broader discourse information to interpretation. Using an experimental paradigm largely similar to the current one, Luce and colleagues (2018) present an experiment comparing the interpretation of VPE and the verbal anaphor *do that* in complex discourse contexts. *Do that* has long been considered a 'deep' anaphor that can relatively freely receive its interpretation from the broader discourse context as well as from a linguistic antecedent (Hankamer & Sag 1976, Merchant 2004, Miller & Pullum 2014). The experimental results indicated that while both VPE and *do that* consider antecedent and broader discourse information, the effect of discourse information is stronger for *do that* than for VPE. Together, these observations suggest that the exact division of labor between a narrow, identity-driven interpretive module and a broader one integrating broader contextual information is partly dependent on the form of the linguistic expression. Different verbal anaphors, such as VPE and *do that*, consult the linguistic and nonlinguistic discourse in different ways. This is perhaps unsurprising, as such form-specific interpretive biases that are known to be in play for nominal anaphors (e.g. Gundel et al. 1993). For instance, Kehler and Rohde (2013) made similar observations for pronoun resolution, suggesting that interpretation is guided both by narrow grammatical correspondences between pronouns and their antecedents as well as by broader listener reasoning about the likely meaning of the discourse.

Finally, we acknowledge that there has been a fruitful line of work on the effects of both discourse coherence relations and information structural considerations on the acceptability of ellipsis sentences with formal mismatches and, by likely extension, the possible interpretations for ellipsis sites (Kehler 2000, 2002; Kertz 2008, 2013; Poppels & Kehler 2019). The present study featured a relatively restricted manipulation in the information available in the discourse context, and did not manipulate either the coherence relations at play in the discourse or its information structure. We see our results as broadly compatible with this line of work investigating discourse effects on ellipsis resolution. Future work combining the present experimental paradigm with manipulation of a wider variety of discourse factors may further clarify how grammatical constraints and broader discourse considerations interact during VPE interpretation.

6. CONCLUSION. Using a novel experimental paradigm, the current study examined the interpretation of verb phrase ellipsis in context. We showed that consulting broad discourse information is a regular strategy for ellipsis resolution, instead of a secondary accommodation or repair strategy. At the same time, an identity relationship with the linguistic antecedent remains the dominant, albeit defeasible, factor in determining the interpretation of a VPE site. These findings help to constrain the set of possible mechanisms according to which information from linguistic antecedents and from representations of discourse at large is combined in the interpretation of VPE utterances. We also anticipate that similar experimental paradigms may be used in the future to further investigate the division of labor between different levels of linguistic representation.

REFERENCES

ARREGUI, ANA; CHARLES CLIFTON, JR.; LYN FRAZIER; and KEIR MOULTON. 2006. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language* 55.232-46.

BATES, DOUGLAS; MARTIN MAECHLER; BEN BOLKER; AND STEVEN WALKER. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67.1-48. Online: https://www.jstatsoft.org/article/view/v067i01

BERGEN, LEON, and NOAH D. GOODMAN. 2015. The strategic use of noise in pragmatic reasoning. *Topics in Cognitive Science* 7.336-50.

BÜRKNER, PAUL-CHRISTIAN. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80.1-28. Online: https://www.jstatsoft.org/article/view/v080i01

DALRYMPLE, MARY; STUART M. SHIEBER; AND FERNANDO C. N. PEREIRA. 1991. Ellipsis and higher-order unification. *Linguistics and Philosophy* 14.399-452.

DRUMMOND, ALEX. 2020. *Ibex Farm manual*. Online: https://github.com/addrummond/ibex/blob/master/docs/manual.md

FOX, DANNY. 2000. *Economy and semantic interpretation*. Cambridge, MA: The MIT Press.

FRANKE, MICHAEL. *Signal to act: Game theory in pragmatics*. Amsterdam: Universiteit van Amsterdam dissertation.

FUTRELL, RICHARD, and ROGER LEVY. 2017. Noisy-context surprisal as a human sentence processing cost model. *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics*, 15.688-98.

GEIGER, JEFFREY, and MING XIANG. 2017. "Context can!": Contextual accommodation in exophoric and anaphoric verb phrase ellipses. *Proceedings of the North East Linguistic Society*, 47.35-44 (Volume 2).

GIBSON, EDWARD; LEON BERGEN; and STEVEN T. PIANTADOSI. 2013. Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences* 110.8051-56. Online: https://www.pnas.org/content/110/20/8051

GOODMAN, NOAH D., and ANDREAS STUHLMÜLLER. 2013. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science* 5.173-84.

GRICE, H. PAUL. 1975. Logic and conversation. *Studies in syntax and semantics III: Speech acts*, ed. by Peter Cole and Jerry L. Morgan, 183-98. New York: Academic Press.

GUNDEL, JEANETTE K.; NANCY HEDBERG; and RON ZACHARSKI. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69.274-304.

HAÏK, ISABELLE. 1987. Bound VPs that need to be. *Linguistics and Philosophy* 10.503-30.

HANKAMER, JORGE, and IVAN SAG. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7.391-428.

HARDT, DANIEL. 1993. *Verb phrase ellipsis: Form, meaning, and processing*. Philadelphia, PA: University of Pennsylvania dissertation.

JACOBSON, PAULINE. 2003. Binding without pronouns (and pronouns without binding). *Resource-sensitivity, binding, and anaphora*, ed. by Geert-Jan M. Kruijff and Richard T. Oehrle, 57-96. Dordrecht: Kluwer Academic Publishers.

JÄGER, GERHARD. 2011. Game-theoretical pragmatics. *Handbook of logic and language*, ed. by Johan F. van Benthem and Alice ter Meulen, 467-91. London: Elsevier, 2nd edition.

KEHLER, ANDREW. 2000. Coherence and the resolution of ellipsis. *Linguistics and Philosophy* 23.533-75.

KEHLER, ANDREW. 2002. *Coherence in discourse*. Stanford, CA: CSLI Publications.

KEHLER, ANDREW, and HANNAH ROHDE. 2013. A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics* 39.1-37.

KERTZ, LAURA. 2008. Focus structure and acceptability in verb phrase ellipsis. *Proceedings of the West Coast Conference on Formal Linguistics*, 27.283-91.

KERTZ, LAURA. 2013. Verb phrase ellipsis: The view from information structure. *Language* 89.390-428.

KIM, CHRISTINA S.; GREGORY M. KOBELE; JEFFREY T. RUNNER; and JOHN T. HALE. 2011. The acceptability cline in VP ellipsis. *Syntax* 14.318-54.

LASSITER, DANIEL, and NOAH D. GOODMAN. 2017. Adjectival vagueness in a Bayesian model of interpretation. *Synthese* 194.3801-36.

LEVY, ROGER. 2008. A noisy-channel model of rational human sentence comprehension under uncertain input. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 234-43. Stroudsburg, PA: Association for Computational Linguistics. Online: https://dl.acm.org/doi/pdf/10.5555/1613715.1613749

LUCE, KANAN B.; JEFFREY GEIGER; CHRISTOPHER KENNEDY; and MING XIANG. 2018. Interpretations of VP anaphora through reference to salient events. *Proceedings of the Linguistic Society of America* 3.1-13 (Issue 38). Online: https://journals.linguisticsociety.org/proceedings/index.php/PLSA/article/view/4326/3930

MERCHANT, JASON. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27.661-738.

MERCHANT, JASON. 2013. Voice and ellipsis. *Linguistic Inquiry* 44.77-108.

MERCHANT, JASON. 2019. Ellipsis: A survey of analytical approaches. *The Oxford handbook of ellipsis*, ed. by Jeroen van Craenenbroeck and Tanja Temmerman, 19-45. Oxford: Oxford University Press.

MILLER, PHILIP, and GEOFFREY K. PULLUM. 2014. Exophoric VP ellipsis. *The core and the periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag*, ed. by Philip Hofmeister and Elisabeth Norcliffe, 5-32. Stanford, CA: CSLI Publications.

POPPELS, TILL, and ANDREW KEHLER. 2019. Reconsidering asymmetries in voice-mismatched verb phrase ellipsis. *Glossa* 4.60. Online: https://www.glossa-journal.org/articles/10.5334/gjgl.738/

RYSKIN, RACHEL; RICHARD FUTRELL; SWATHI KIRAN; and EDWARD GIBSON. 2018. Comprehenders model the nature of noise in the environment. *Cognition* 181.141-50.

SAG, IVAN. *Deletion and logical form*. Cambridge, MA: MIT dissertation.

SAG, IVAN, and JORGE HANKAMER. 1984. Toward a theory of anaphoric processing. *Linguistics and Philosophy* 7.325-45.

SCHACHTER, PAUL. 1977. Does she or doesn't she? *Linguistic Inquiry* 8.763-67.

THOMS, GARY. 2015. Syntactic identity, parallelism, and accommodated antecedents. *Lingua* 166.172-98.

VAN CRAENENBROECK, JEROEN. 2012. Ellipsis, identity, and accommodation. Brussels: KU Leuven, MS. Online: https://static1.squarespace.com/static/5217e223e4b090faa01f8f2d/t/54081f1be4b0371c21188146/1409818395988/paper+ellipsis+and+accommodation.pdf

WASOW, THOMAS. 1972. *Anaphoric relations in English*. Cambridge, MA: MIT dissertation.

WEBBER, BONNIE. 1978. *A formal approach to discourse anaphora*. Cambridge, MA: Harvard University dissertation.

WILLIAMS, EDWIN. 1977. Discourse and logical form. *Linguistic Inquiry* 8.101-39.

NOTES

[1] For a general review of different types of evidence regarding the syntactic and semantic identity constraints, see Merchant 2019.

[2] glmer model specification: Response ~ Context * Antecedent + (1|Participant) + (1+Context+Antecedent|Item), family=binomial. To attain model convergence, we selected random participant effects for greater effect reduction due to the sparse nature of by-participant responses.

[3] glmer model specification: Response ~ Context + (1|Participant) + (1+Context|Item), family=binomial.

[4] glmer model specification: Response ~ 1 + (1|Item), family=binomial. Note that no random effect was included for participants because each participant saw each condition combination at most one time.

[5] Analysis of the data with this mental state included did not yield qualitatively different results.

[6] The contrast coding for the context variable is (-1, -1, 2) for the unavailable, available, and salient levels, respectively. For the antecedent variable, the coding is (-1, 2, -1) for the exophoric, unmodified, and modified levels, respectively. The experiment variable is treatment coded as (0, 1) with Experiment 1 as the baseline. All models reported in section 4.3 shared the same contrast coding.

[7] The R code for the brms model is the following.

```
prior = c(set_prior("normal(0,1.5)", class="b"),
        set_prior("normal(0,1)", class="Intercept"),
        set_prior("lkj(2)", class="cor"))
model = brm(Response ~ Context * Experiment + (1+Context*Experiment|Item) +
(1+Context|Participant), data=data, family=Bernoulli(link=logit), prior=prior, iter=6000, control
= list(adapt_delta = 0.95), seed=500)
```

Since experiment is a between-participant variable, it was not included for the by-participant random effects.

[8] brms model specification: Response ~ Antecedent * Experiment + (1+Antecedent*Experiment|Item) + (1+Antecedent|Participant).

[9] brms model specification: Response ~ Experiment + (1+Experiment|Item).

[10] brms model specification: Response ~ Antecedent * Context * Experiment + (1+Antecedent*Context*Experiment|Item) + (1+Antecedent*Context|Participant).