

A probabilistic account of verb phrase ellipsis interpretation in context

Jeffrey Geiger¹ & Ming Xiang²

^{1,2}Department of Linguistics
University of Chicago
1115 E 58th Street
Rosenwald Hall #224
Chicago, IL 60637-1511
United States

¹Corresponding author
jgeiger@uchicago.edu

Abstract

The interaction between linguistic structure and broader representations of discourse during the interpretation of context-sensitive linguistic expressions is poorly understood. One area where this interaction is brought to bear is the interpretation of verb phrase ellipsis. Previous accounts have proposed that prior linguistic content and the discourse availability of the competing interpretations are both relevant to the interpretation of verb phrase ellipsis, but little has been done to determine the relative contributions of information from these two sources. We investigated the interpretation of verb phrase ellipsis in contexts where the linguistic antecedent and the broader discourse context make different information available for inclusion in the interpretation of the ellipsis site. The results indicate that neither a strict identity strategy nor a strict discourse strategy is sufficient to account for the distribution of available interpretations. Our experimental findings suggest that the distribution of available interpretations is sensitive to features of the broader discourse context, even when there is a viable linguistic antecedent, but that the available interpretations are also constrained by considerations of identity with the antecedent. Following recent development in probabilistic pragmatic models, we propose a hybrid interpretation model that probabilistically combines information from both the linguistic antecedent and the discourse. The hybrid model substantially outperforms two other models we constructed that rely exclusively on either the linguistic antecedent or the broader discourse. The experimental and modeling results presented in this paper therefore make contributions toward better understanding both the connections and the boundaries between the linguistic antecedent and mental models of the context of utterance.

Keywords

verb phrase ellipsis; meaning in context; syntax-discourse interface; discourse comprehension; probabilistic pragmatics

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. Consider the example in (1):

(1) A: Don't!

The sentence in (1) is an example of an elliptical expression. Informally speaking, the elliptical reply in (1) is inherently “incomplete”, since the verb phrase is missing (hence the term verb phrase ellipsis, or VPE). 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, demonstrated in (2), is that VPE can be interpreted using the content of an antecedent verb phrase that was uttered in the recent linguistic context.

(2) B: I'm going to move the candle.

A: Don't!

In (2), it seems fairly intuitive that A intends for their utterance to be interpreted as meaning *Don't move the candle!* This stands in contrast to the second observation, which is that under certain conditions, VPE can be interpreted without a linguistic antecedent in a sufficiently informative context, as shown in (3).

(3) [B moves their hand dangerously close to a lit candle]

A: Don't!

In the case of (3), one reasonable interpretation of A's utterance in context is *Don't touch the flame!*

The exchanges in (2) and (3) raise fundamental questions about how VPE is interpreted. The main question is whether interpretation under identity with a linguistic antecedent, as in (2), and interpretation with respect to a nonlinguistic representation of the discourse context, as in (3), can be analyzed as representing a single, unified interpretive mechanism. If they cannot,

then the natural follow-up is to examine the relative strength of each strategy in determining the interpretation of a VPE utterance and the mechanism by which the predictions of each strategy are combined.

In this paper, we present novel empirical and modeling evidence to support a hybrid probabilistic account of VPE interpretation. Our experimental findings indicate that the interpretation mechanism for verb phrase ellipsis uttered in context cannot be described as depending only on prior linguistic information or only on the prior discourse salience of the competing interpretations. Furthermore, by comparing a number of probabilistic models based on the experimental data, we show that the model that best accounts for the data is one that combines both the linguistic information in the antecedent and contextual information from the broader discourse.

2. Background

The two observations presented above recall common topics of discussion from the literature on the broader family of elliptical constructions. In (2), the ellipsis site and the antecedent verb phrase stand in a relationship of *identity* - that is, the ellipsis site is interpreted as though it contains content identical to the antecedent phrase. There is a large body of literature discussing the role of antecedent-ellipsis site identity in licensing ellipsis - that is, in determining when elliptical constructions can felicitously be used.

A key battleground for the various identity-driven accounts of licensing is cases of formal antecedent-ellipsis mismatch. In some cases, the pronounced part of the clause containing ellipsis forces a particular structural parse of the ellipsis site - for instance, that it has active or passive voice. Critically, these inferred parses can either be consistent (match) or inconsistent (mismatch) with the parse of the linguistic antecedent. In some cases, formal mismatches between the antecedent and the ellipsis site are judged as acceptable, but in other cases, such mismatches rule elliptical structures out (Bresnan, 1971; Warner, 1985; Hardt, 1993; Fiengo and May, 1994; Kehler, 2002; Merchant, 2008, i.a.).

A principal debate among proponents of identity-driven accounts concerns whether the relationship of identity between the antecedent and the ellipsis site is predicated on syntactic or semantic grounds. Requiring exact syntactic identity between the antecedent and the ellipsis site readily accounts for the cases in which sentences with formal antecedent-ellipsis mismatches are judged as infelicitous (Sag, 1976; Hankamer and Sag, 1976; Rooth, 1992; Fiengo and May, 1994, i.a.). Likewise, semantic identity accounts, which allow for formal mismatches as long as the antecedent

and ellipsis site have the same interpretation, correctly generate those mismatch sentences that are judged acceptable (Dalrymple et al., 1991; Hardt, 1993; Ginzburg and Sag, 2000; Merchant, 2001, i.a.).

The empirical landscape is further complicated by the observation that certain types mismatch between the linguistic antecedent and the ellipsis site cause varying degrees of degradation of the sentence’s acceptability. Work dealing with the possibility of antecedent-ellipsis mismatch and the acceptability cline has led to proposals allowing for the accommodation (Fox, 1999; van Craenenbroeck, 2013; Thoms, 2015) or construction via repair (Arregui et al., 2006) of a new linguistic antecedent, deriving the acceptability cline from independent processing heuristics (Kim et al., 2011), or linking the acceptability cline to differences in information or discourse structure (Kehler, 2002; Kertz, 2013).

In the current paper, our primary goal is not to distinguish between the classes of identity account. Since all of these proposals primarily focus on the concept of linguistic identity (syntactic or semantic) with an antecedent to model the interpretation of a verb phrase ellipsis site, for the purpose of the current study, we will refer to this class of account as the *Identity* proposals. While the work cited above deals principally with ellipsis licensing rather than interpretation, it seems straightforward to propose that there is an identity requirement on interpretation, as well. Thus, we will entertain the hypothesis that the VPE interpretation strategy, or one component of it, involves calculating the meaning of an ellipsis site under exact identity with the linguistic antecedent, and we will call this approach the *Identity strategy*.

Beyond the identity literature, the availability of the *Don’t touch the flame reading* in (3) recalls a smaller body of literature discussing the possibility of ellipsis in the absence of a viable linguistic antecedent. With the appropriate context, it has been observed that VPE-like constructions can be used in the absence of a linguistic antecedent (so-called *antecedentless* or *exophoric* VPE). Examples are shown in (4) and (5).

(4) [*John comes to the table where Mary is sitting and makes as if to take one of the spare chairs there.*]

John: May I?

Mary: Please do. (Schachter, 1977)

(5) [*The aisles at the Lakewood Wal-Mart are surprisingly packed at 11 p.m.*]

Vanessa: Can we? Can we?

[*Vanessa tugs at her mother, pointing to a rack of “Lady and the Tramp” DVDs. Diaz shrugs. OK.*] (Miller and Pullum, 2013)

The utterances in (4) are clearly meant to be interpreted something like *May I take this chair* and *Please do take this chair*. Likewise, the question *Can we* in (5) is straightforwardly interpreted as meaning something like *Can we buy a “Lady and the Tramp” DVD*. This is in spite of the fact that there is no antecedent VP available in the linguistic context that supplies a meaning even close to the one ascribed to the ellipsis site. Evidently, the information in the nonlinguistic context is sufficient for listeners to relatively easily construct a meaning for the ellipsis site.

While it has been suggested that these cases are somehow exceptional (Hankamer, 1978), the evidence indicates that antecedentless ellipsis can be used productively, although it is rare (Schachter, 1977; Merchant, 2004; Miller and Pullum, 2013). These antecedentless cases are problematic for Identity accounts of VPE licensing, which do not predict that VPE should be acceptable in the absence of a linguistic antecedent. Likewise, it is not immediately apparent how the Identity strategy of interpretation could account for the relative ease with which meanings are assigned to these utterances.

One possibility for handling such cases is to allow a more general representation of the discourse to play a larger and more direct role in determining the licensing and interpretation of VPE utterances. A number of analyses have taken up this approach, and we will refer to this class of proposals as the *Discourse strategy* of VPE interpretation. While the exact implementations vary, the general approach of the discourse strategy is to resolve the ellipsis site by allowing it to directly access the discourse at large, with listeners inferring that a VPE utterance should address something that is salient in the discourse (Schachter, 1977; Webber, 1979; Hardt, 1992; Kehler, 1993; Miller and Pullum, 2013; Poppels and Kehler, 2018, 2019).

Under the strongest version of the Discourse strategy, VPE interpretation should be guided entirely by the prior discourse status of the competing meanings, and linguistic identity plays no overt role. Miller and Pullum (2013) present a VPE licensing account that suggests an interpretation strategy of this sort. In their analysis, ellipsis is licensed exactly when a set of contrasting alternative propositions is made salient in the discourse, and the VPE utterance picks out one alternative as correct or otherwise preferable. For example, in (5), Miller & Pullum suggest that the discourse makes one contrast set of alternative propositions highly salient, namely the contrast between *buy a DVD* and *not buy a DVD*. Thus, a VPE reply that clearly picks out one of the alternatives, like *Can we*, is licensed. By extension, then, we could presume that a VPE utterance should be

interpreted by retrieving the salient alternative set that licensed the utterance and selecting the proposition with which the utterance is compatible.

The discussion of antecedentless VPE raises several possibilities for uniting our two fundamental interpretation observations. The first possibility is that VPE interpretation always proceeds under strict identity, and that antecedentless cases, despite their surface structure, are not actually VPE at all. Merchant (2004) takes this tack, proposing that when faced with apparent antecedentless VPE, listeners infer that the elided material is actually an anaphor that can freely receive its interpretation from broader discourse information, like *do that*. We will discuss the merits of such an approach further in the general discussion.

Likewise, it is possible that the two observations can be united under the auspices of interpretation with respect to the discourse context at large. Miller & Pullum claim that the apparent requirement for identity between an ellipsis site and an overt antecedent is an artifact of the extreme salience of overt linguistic information over nonlinguistic information, but that a sufficiently supportive context can license VPE even in the absence of an antecedent. Under the extended interpretation account, the interpretation of a VP ellipsis site would be driven entirely by the prior discourse availability of each of the meanings under consideration - that is, the search for a VPE interpretation is a search for the most salient contrast set of propositions in the discourse at large. What was previously analyzed as interpretation under identity with a linguistic antecedent might actually be purely discourse-driven interpretation, with the highly salient linguistic antecedent simply dominating other elements, particularly nonlinguistic ones, of the broader discourse immediately before the use of the VPE utterance.

Alternately, it may be the case that the two interpretive strategies make completely independent contributions to the meanings listeners infer for VPE sites in complex discourse contexts. If this is correct, the two strategies should not be collapsed, and further work is licensed investigating the exact mechanism by which the predictions of the two strategies are synthesized during the task of VPE interpretation.

The Identity strategy and the Discourse strategy face different challenges in terms of empirical coverage. In general, the Identity strategy undergenerates elliptical sentences, as it does not predict that VPE should be interpretable in antecedentless form. The Discourse strategy handles antecedentless ellipsis well, but it overgenerates with respect to formal mismatches between the ellipsis site and the antecedent; such mismatches often lead to acceptability degradation despite having an interpretation that is completely coherent in the general discourse context. Proposals

endorsing one of these strategies tend to highlight the respective empirical advantages, paying less attention to a systematic investigation of the interaction between the two strategies.

The empirical goal of the current study is to fill in this gap. In particular, in the two experiments we will present below, we simultaneously vary the information available in the linguistic antecedent and from the broader discourse context. This allows us to observe how the two sources of information compete with each other for recruitment during interpretation and whether one strategy (Identity or Discourse) is subordinate to the other. After we present the experimental data, to formalize an explanation of the empirical observations, we consider three probabilistic models of VPE interpretation. These models transform the prior probabilities of the competing VPE interpretations into predicted posterior interpretation distributions according to the Identity strategy, the Discourse strategy, or a combination of the two. Model comparison indicates that the hybrid model that combines both strategies substantially outperforms the other two.

3. Experiment 1: VPE interpretation in context

The goal of Experiment 1 is to systematically assess the empirical coverage of the Identity strategy and the Discourse strategy. Previous work on the Identity strategy paid little attention to information available in the broader discourse, including the nonlinguistic context. On the other hand, the Discourse strategy, especially proposals motivated by the antecedentless VPE cases, focused primarily on the facilitating role of the broader discourse context and attempted to reduce any effect of a linguistic antecedent to a more general discourse effect. 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. If participants' likelihood to infer a numeral in the ellipsis site is modulated only by the manipulation of number information in the linguistic antecedent, the Identity strategy of interpretation is supported. If ellipsis interpretation tracks the prior salience of the number information in the discourse at large, considering both antecedent and non-antecedent information, then the Discourse strategy is supported.

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 subjects could select a choice by clicking on it or by pressing a number key on their keyboard. A sample experimental item is shown in Figure 1.



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 item

The experiment featured a 3 x 3 (comic strip context x 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.

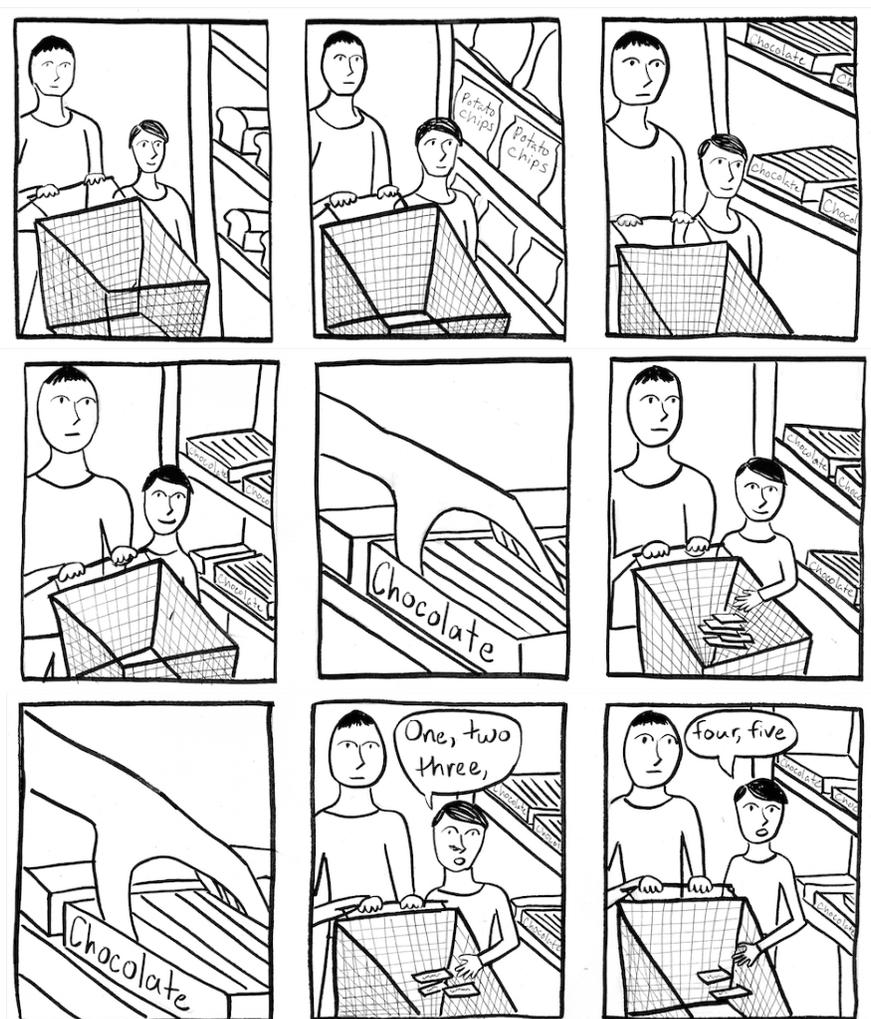


Figure 2: Top to bottom: Unavailable, Available, and Salient comic strip contexts for a sample Experiment 1 item

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 an utterance 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 is attributed to the son and is *I want to buy candy bars!* 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 (2013) 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, e.g., *We can't buy (any) candy bars*. The Scalar interpretation represents interpretation of the ellipsis site with a numeral, e.g., *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	[Son stands near candy bars.] Father: We can't.	[Son stands near candy bars.] Son: I want to buy candy bars! Father: We can't.	[Son stands near candy bars.] Son: I want to buy five candy bars! Father: We can't.
Available Context	[Son takes five candy bars together.] Father: We can't.	[Son takes five candy bars together.] Son: I want to buy candy bars! Father: We can't.	[Son takes five candy bars together.] Son: I want to buy five candy bars! Father: We can't.
Salient Context	[Son counts out five candy bars.] Father: We can't.	[Son counts out five candy bars.] Son: I want to buy candy bars! Father: We can't.	[Son counts out five candy bars.] Son: I want to buy five candy bars! Father: We can't.
<p><i>Prompt:</i> Based on the scenario above, do you think it is more likely that the father meant:</p> <p><i>Polar:</i> We can't buy any candy bars.</p> <p><i>Scalar:</i> We can't buy five candy bars, but maybe we could buy fewer.</p>			

Table 1: Conditions and interpretation prompts for a sample scenario 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. Subjects provided informed consent and completed a demographic questionnaire and two unrelated practice items before beginning the experiment itself.

Subjects 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 subjects 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 numeral. The VPE replies in all the scenarios 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 one item from each of the six critical scenarios, and 10 filler items that involved VPE interpretation, but without the numeral manipulation. For each participant, the 6 critical trials were drawn from 6 different conditions.

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 indicate themselves as native English speakers in the demographic survey. The number of observations of each scenario and each condition was roughly evenly distributed.

3.5. Results

The proportions of Polar and Scalar responses by condition are shown in Figure 3.

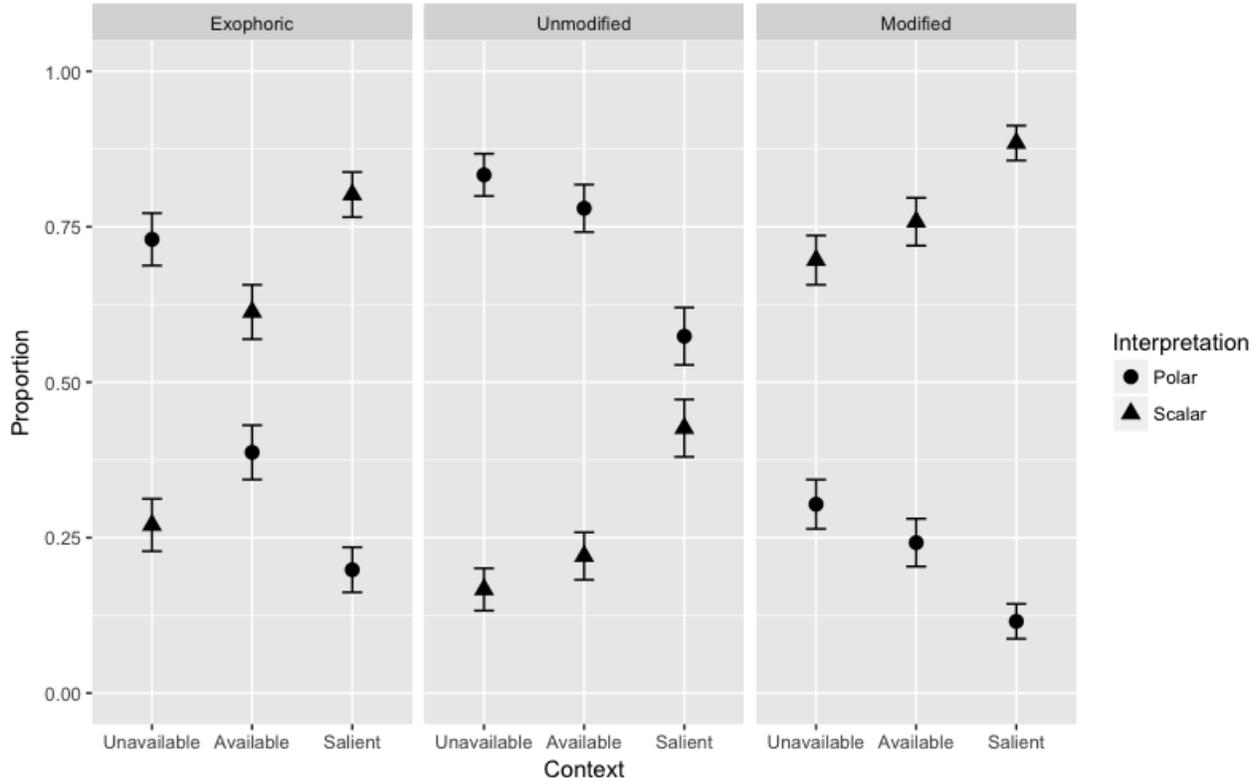


Figure 3: Experiment 1 results plotted as proportions. The plot is split horizontally by antecedent, with context plotted along the x-axis in each frame. Error bars: standard error.

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.

3.6. Analysis

The response data were coded with Polar responses as a success and Scalar responses as a failure. Two logistic mixed effects regression models were constructed. The first predicted the

response data with fixed effects for comic strip context, antecedent, and their interaction, and with random effects for subject and item. The second had fixed effects for comic strip context and antecedent, and random effects for subject and item, but no interaction between 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.

For each antecedent condition, we performed two types of analyses. First, 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 mean proportion of the Polar response at each level of the context was compared to the mean 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 ($p > .3$), 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 ($p > .2$), 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 binomial test was performed in each condition testing whether the proportion of Polar responses was significantly different (higher or lower) from 0.5. The proportion was significantly different from 0.5 at the $p = .001$ level in every condition in Figure 3 except in the Exophoric/Available condition, where the difference from 0.5 was significant at the $p = .01$ level, and the Unmodified/Salient condition, where the proportion was not significantly different from 0.5 ($p > .1$).

3.7. Discussion

Experiment 1 tested how VPE is interpreted in a rich discourse context, where material from both the linguistic discourse context and the general discourse context - linguistic and nonlinguistic - are available during interpretation. In the Exophoric conditions, it seems straightforward that

under both an Identity and a Discourse account, VPE interpretation should proceed according to the availability of the competing interpretations in the general discourse. Since there is no viable linguistic antecedent, an Identity account must include some “backup” mechanism for interpretation according to discourse availability in these situations.

When there is a linguistic antecedent, in the Unmodified and Modified conditions, the interpretation mechanism differs according to the Identity and Discourse strategies. With a viable linguistic antecedent, the Identity strategy mandates that VPE be interpreted under identity with the antecedent. This means that in the Unmodified conditions, the VPE reply should be interpreted as though the elided VP is *buy candy bars*. Since this VP is negated in the reply, the only possible interpretation is that the father means no candy bars can be bought (i.e., *we can't buy candy bars*). Thus, the Identity account predicts that only the Polar interpretation should be available in the Unmodified conditions. By contrast, the Discourse account predicts that the Scalar interpretation could also be available if the corresponding proposition were at least somewhat salient in the prior discourse context.

In the Modified conditions, the Identity account predicts that the elided VP should be interpreted as *buy five candy bars*. Since this VP is negated (i.e., *we can't buy five candy bars*), two interpretations are available: either no candy bars can be bought, or fewer than five candy bars can be bought. Thus, both the Polar and Scalar interpretations are available according to the Identity strategy, although it is not immediately apparent what governs the relative strength of the two interpretations. Again, the Discourse account predicts that both interpretations should be available in proportion to their prior salience in the discourse.

Thus, the most critical difference in predictions between the Identity and Discourse accounts for Experiment 1 are in the Unmodified antecedent conditions. In these conditions, the Identity account predicts that only the antecedent-identical Polar interpretation should be available, while the Discourse account allows for the Scalar interpretation to be available in proportion to its prior salience.

Turning to the results, in the Exophoric conditions, VPE interpretation was largely a function of the intended availability of number information in the comic strip context. With an Unavailable context, the Polar interpretation was significantly preferred, but as number information became more available in the context, the Scalar interpretation gained strength and was preferred with both the Available and Salient contexts. This pattern of results 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 (e.g., Miller and Pullum, 2013), which suggests that in the absence of a linguistic antecedent, VPE should be interpreted with respect to the information generally available in the discourse context.

The conditions with antecedents likewise show effects of context. In the Unavailable context, where there was little numeral information in the nonlinguistic context, each antecedent type corresponded to a clearly preferred interpretation - the antecedent-identical Polar interpretation for Unmodified antecedents, and the Scalar interpretation for Modified antecedents. Switching the context to Available did not significantly shift the interpretations in either case, although the results for both antecedent types suggest a possible non-significant trend in favor of the Scalar interpretation. With a Salient context, however, we observe a clear effect of the comic strip context. In the Unmodified condition, the antecedent-nonidentical interpretation, the Scalar interpretation, received a significant boost and was treated as approximately equally likely as the antecedent-identical interpretation. In the Modified condition, 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 were reproduced in another study using a 1-7 Likert scale for rating the two VPE interpretations separately (Geiger and Xiang, 2017; Luce et al., 2018). Altogether, the results of Experiment 1 are problematic for a strict formulation of the Identity strategy of VPE interpretation. According to this strategy, 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.

The results from Experiment 1 are potentially compatible with the Discourse strategy of VPE interpretation. According to this strategy, VPE is interpreted according to the prior discourse status of each competing interpretation. It is possible that the proportions for each interpretation in Experiment 1 map closely to the proportion of times a Polar (e.g., *buy candy bars*) versus Scalar (e.g., *buy five candy bars*) proposition would be inferred as being most salient in the discourse prior to the VPE reply. In order to conclude that the Discourse strategy is correct, however, we must empirically verify that this is the case - that is, that there is a close mapping between the Experiment 1 interpretation proportions and the propositions that are available in each discourse

prior to the VPE utterance. Experiment 2 was conducted to empirically assess the prior probability of each interpretation before the VPE reply.

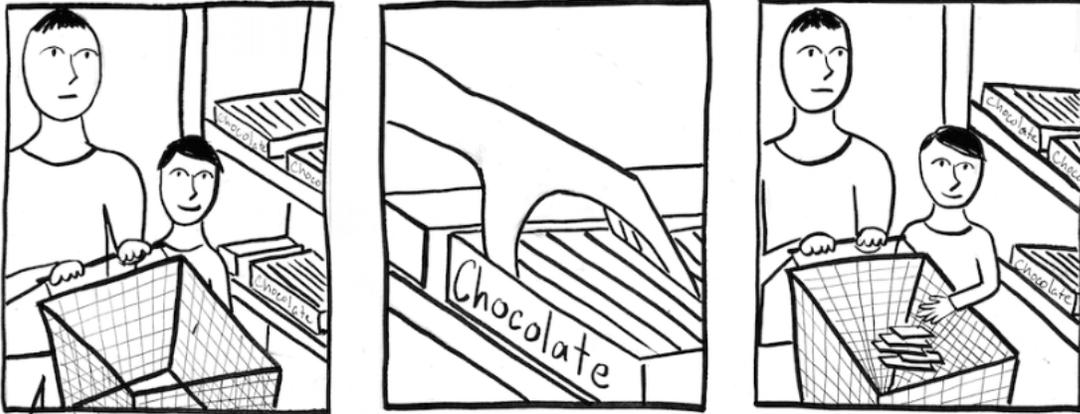
4. Experiment 2: Empirically measuring priors

In Experiment 2, we assess the discourse salience of the competing propositions in the Experiment 1 items prior to the VPE utterance. Since the VPE utterance from the second character in Experiment 1 was always a denial of a request or proposal, in Experiment 2 we probed the likelihood of each possible request or proposal the first character could have intended. Based on these results, we can infer the prior discourse availability of each of the competing VPE interpretations from Experiment 1. The purpose of this experiment is twofold. First, the results from Experiment 2 can be compared to those from Experiment 1 to determine whether the Discourse strategy alone is a valid account of VPE interpretation in context. Second, the Experiment 2 results will be used below as the prior probabilities of the competing interpretations in our probabilistic models of VPE interpretation.

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 mind states were most likely for the character that utters the antecedent. The mind 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. A sample item from Experiment 2 is shown in Figure 4.

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: Layout of Experiment 2 item

The main goal of Experiment 2 was to assess what a listener would perceive as the most salient “issue” in the discourse before the VPE utterance was uttered. Under the Discourse-based proposal for VPE interpretation, there is a natural connection between what is at-issue in the discourse context and the interpretation a VPE utterance receives. That is, if a listener decides, after hearing the antecedent sentence, that what is at-issue is buying candy bars, regardless of number (Indifferent), then the VPE reply *We can't* should indicate that no candy bars can be purchased, which is the Polar interpretation of VPE. If the listener decides that buying a specific number of candy bars is at issue (Specific), then *We can't* should mean that that number of candy bars can't be purchased, which is the Scalar interpretation. For the purposes of this experiment, we assume that the probability of inferring the Indifferent mind state before the VPE utterance corresponds to the prior probability of the Polar interpretation, and the probability of inferring the Specific mind state corresponds to the prior probability of the Scalar interpretation.

The third choice, described above as the Zero mind state, was included in Experiment 2 for the

sake of naturalness, as it might be the most natural choice in the Unavailable/Exophoric 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. Since the main goal of Experiment 2 is to help us understand and model the VPE interpretations, we will not discuss participants' choice of the Zero mind state in the data report below. We note here that the Zero mind state was chosen only 8.1% of the time overall, with 41.4% of all Zero responses falling in the Unavailable/Exophoric condition.

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

The proportions of Indifferent and Specific responses, calculated after discarding the Zero responses, are plotted in Figure 5.

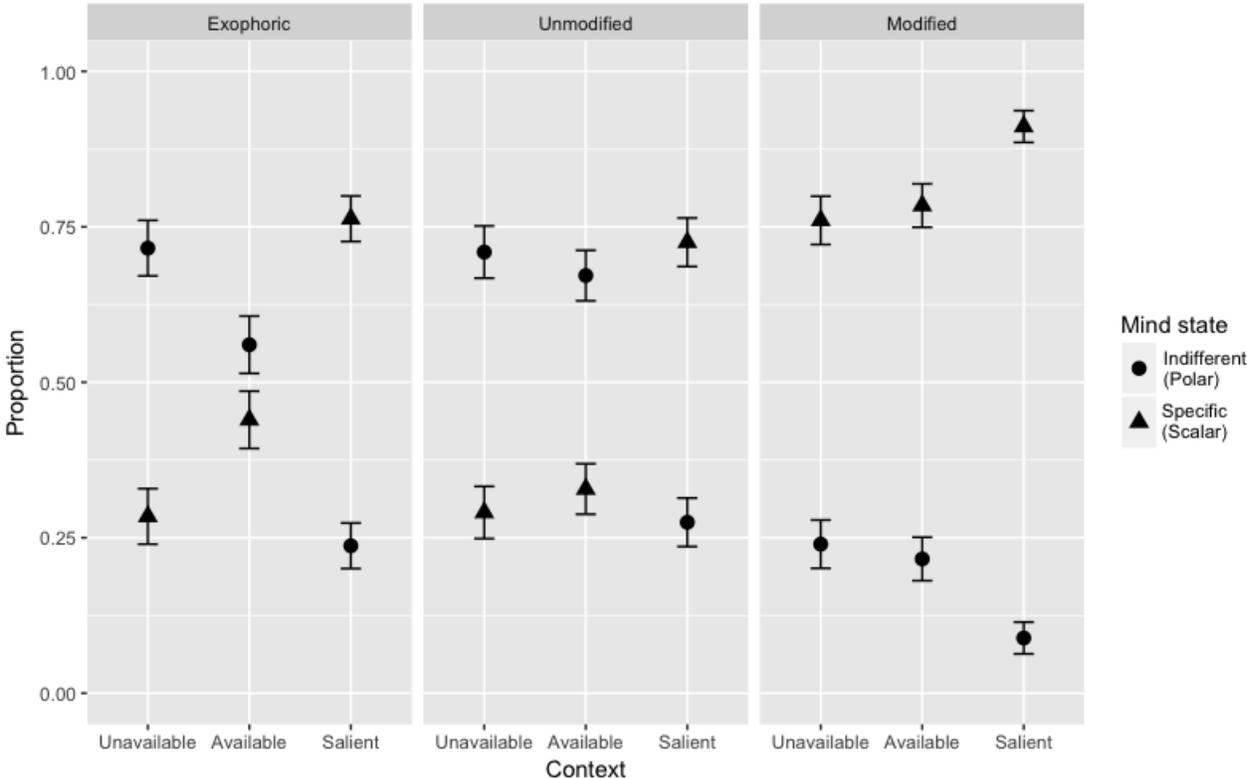


Figure 5: Experiment 2 Indifferent/Specific response proportions. Error bars: Standard error.

4.4. Analysis

As in Experiment 1, a binomial test was performed in each condition to determine whether the proportion of Indifferent responses was significantly different from a random choice between the two mind states, represented by the proportion 0.5. The proportion of Indifferent responses was significantly different from 0.5 at the $p=.001$ level in every condition except Available/Exophoric, where the proportion was not significantly different from 0.5 ($p>.2$).

4.5. Discussion

The goal of Experiment 2 is to estimate what are the most salient “issues”, or propositions supported by the broad discourse context prior to the VPE utterance. Under the Discourse strategy, VPE is always interpreted relative to a discourse-salient issue, so the Discourse strategy therefore predicts that the results from Experiment 2 should be completely parallel to the results from Experiment 1. Qualitatively speaking, when Figure 5 and Figure 3 are compared, this prediction is borne out for most of the conditions. However, one critical condition that strongly challenges the Discourse-only strategy is the Salient/Unmodified condition. In Experiment 1, neither interpretation was preferred significantly above chance level, but in Experiment 2, the Specific mind state was preferred by a substantial margin.

This result indicates that the Discourse strategy is not an adequate account of VPE interpretation in context. According to the Discourse 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. This is clearly not observed in Experiment 1, where the Scalar interpretation is not only not significantly preferred beyond chance levels, but is actually numerically slightly dispreferred. The most obvious explanation is that an overt preference for interpretation under linguistic identity inhibits consideration of the Scalar interpretation of the VPE reply in this condition, despite its superior availability in the discourse, in favor of the antecedent-identical Polar interpretation.

Thus, the results of Experiments 1 and 2, considered together, appear to suggest that neither the Identity strategy nor the Discourse strategy are sufficient to capture the facts of VPE interpretation in context. The Identity strategy, strictly formulated, does not allow for VPE interpretation to be affected by manipulations in the context outside the linguistic antecedent, but exactly such effects were observed even in the presence of a viable antecedent in Experiment 1. The Discourse strategy is also inadequate, as it is obvious that the interpretations of VPE in Experiment 1’s

Salient/Unmodified condition do not correspond straightforwardly to the prior availability, as estimated in Experiment 2, of the different interpretations available in the discourse context at large.

Intuitively, the Identity strategy seems to provide too small of a role for discourse information from outside the antecedent, as it fails to account for the fact that manipulations in the comic strip context affect the distribution of available VPE interpretations. Likewise, the Discourse strategy provides too small of a role for linguistic identity, as identity considerations can strongly inhibit an interpretation that is otherwise strongly available in the discourse. This seems to suggest that the best model of VPE interpretation in context would be one that combines the two strategies, taking into account both identity and discourse considerations. Such a model might be able to capture the observation that contextual manipulations can affect the available VPE interpretations even with a viable antecedent, but that the interpretation distribution is still constrained by linguistic identity. In the next section, we formalize this intuition by demonstrating that a probabilistic model of VPE interpretation in context combining both interpretation strategies significantly outperforms models that only consider one strategy.

5. A probabilistic model of VPE interpretation in context

The results of Experiments 1 and 2 provided initial empirical support for the conclusion that VPE interpretation in context is driven by a combination of linguistic identity and broader discourse considerations. In this section, we formalize this claim by constructing a model of VPE interpretation in context that considers both the linguistic antecedent and the broader discourse, and demonstrate that the model captures the empirical data presented above better than a model that considers only linguistic identity or only the prior discourse status of the competing interpretations.

The models we will construct are directly inspired by work in probabilistic pragmatic reasoning (Bergen and Goodman, 2015; Frank and Goodman, 2012; Franke, 2009; Goodman and Stuhlmüller, 2013; Jäger, 2011; Lassiter and Goodman, 2017). Such approaches are attractive to us for two reasons. First, probabilistic pragmatic frameworks deal overtly with listeners' prior probabilistic beliefs regarding the competing interpretations of an utterance. This feature fits well with the task of modeling VPE interpretations, since we have seen that VPE interpretations are sensitive to the discourse salience of the possible interpretations. Second, probabilistic models provide convenient formal tools for representing a listener's interpretation of an utterance as a probability distribution over a set of possible messages. This is a desirable feature given our experimental results, which

have indicated that VPE interpretation is a probabilistic rather than a deterministic phenomenon.

We constructed three models of VPE interpretation in context: the Identity model, the Discourse model, and the Two-strategy hybrid model. We will explain the three different models in more detail below, but a core component all three models will share, which is also the main feature we adopt from previous probabilistic pragmatic models, is the idea that a listener’s belief of a particular message upon hearing an utterance is jointly determined by whether the target message is compatible with the truth conditions of the utterance and the prior likelihood of the target message. This is represented in Equation 1:

$$P(m|u, c) \propto \llbracket u \rrbracket^{m,c} \cdot P(m|c) \tag{1}$$

In this equation, m refers to the different messages in the set of possible alternative messages M that are under consideration. u represents the utterance that the listener is trying to interpret; in our case, it is always the VPE reply utterance. c represents the discourse context in which the utterance was used, including both nonlinguistic information about the discourse environment and linguistic representations of previous utterances (e.g., antecedents). The first factor estimates whether a message is compatible with the utterance by considering the truth conditions of the sentence. The output of this factor is a truth value, either 1 or 0. Since the relevant utterance for the current purpose is always a VPE reply, the question immediately arises as to how we should define the truth conditions of an elliptical utterance. This will become one of the crucial features that distinguish the three models presented below, and so will be discussed separately for each model. The second factor in the equation above represents the prior probability of each possible message.

To preview the upcoming discussion, under the Identity model, the distribution of VPE interpretations is determined entirely according to identity with a linguistic antecedent, modulo an allowance for noisy behavior like inattention. The prior discourse status of different possible interpretations is relevant only when multiple interpretations are compatible with the antecedent VP. Under the Discourse model, the distribution of interpretations is driven entirely by the prior discourse status of each interpretation, modulo noise. The Two-strategy model combines both strategies, allowing for interpretation distributions that are intermediate between those predicted by the other two models.

5.1. Identity model

The Identity model uses Equation 2 to calculate the probability that a hearer of the VPE reply will infer a particular interpretation.

$$P(m|u, c) = (1 - \epsilon) \frac{\llbracket u \rrbracket_{\text{identity}}^{m,c} \cdot P(m|c)}{\sum_{m' \in M} \llbracket u \rrbracket_{\text{identity}}^{m',c} \cdot P(m'|c)} + \epsilon \frac{1}{|M|} \quad (2)$$

In this equation, m refers to the different messages in the set of alternative messages M that are under consideration. Here, the possible messages are the Polar or Scalar interpretations. u is the VPE reply utterance. c represents the discourse context prior to the VPE utterance, which is one of the 9 comic strip context and antecedent combinations from the experiments.

The fraction multiplied by $(1 - \epsilon)$ is the core part of the Identity model. The term $\llbracket u \rrbracket_{\text{identity}}^{m,c}$ estimates whether the target message is compatible with the VPE reply in the discourse context c when the truth condition of the VPE utterance is computed via the Identity strategy. Critically, the truth value of a VPE utterance under the Identity strategy is determined by considering the corresponding full clause obtained by replacing the ellipsis site with the exact VP from the antecedent clause.

When there is an antecedent clause, this procedure is straightforward. In the Unmodified conditions, the antecedent VP is of the form *buy candy bars*, so the VPE utterance *We can't* is evaluated according to the sentence form *We can't buy candy bars*. The VPE utterance will therefore be compatible with the Polar interpretation and incompatible with the Scalar interpretation. For the Modified conditions, the antecedent is of the form *buy five candy bars*, so the VPE utterance *We can't* is evaluated according to the full-VP sentence form *We can't buy five candy bars*, which is semantically compatible with both the Polar and Scalar interpretations.

For the Exophoric conditions, we have a choice of how to set the truth value. On the one hand, the VPE utterance is strictly speaking uninterpretable without an antecedent, and will be incompatible with either of the target messages. On the other hand, we can define VPE as trivially compatible with both messages. The latter option would be an implementation of a slightly weaker version of the Identity strategy, which only claims to be an account for cases in which an antecedent is available, and remains agnostic in cases with no linguistic antecedents. We implement the second option in the current model. A summary of the compatibility ratings according to the Identity strategy are shown in Table 2.

Messages (m)	Antecedent		
	Exophoric [no utterance]	Unmodified <i>I want to buy candy bars!</i>	Modified <i>I want to buy five candy bars!</i>
Polar (can buy 0)	1	1	1
Scalar (can buy <5)	1	0	1

Table 2: Truth values for VPE *We can't* with Polar or Scalar interpretation given three antecedents

The factor $P(m|c)$ in Equation 2 is the prior probability of the interpretation under consideration in the given discourse context, i.e., the comic strip context and antecedent combination. These prior probabilities were empirically estimated in Experiment 2. The denominator of the core fraction is a normalizing constant that ensures the output of the calculation is a proper probability distribution that adds up to 1. Under the Identity strategy, the prior discourse status of different possible interpretations is relevant only when these interpretations are compatible with the VPE utterance (i.e., with a truth value of 1). According to Table 2, under the Unmodified conditions, we would predict the Polar interpretation to be selected 100% of the time, and the Scalar interpretation should never be allowed, since the VPE reply is marked as incompatible with the Scalar interpretation in these conditions.

This is clearly too strong a prediction for participants' behavior in the experiments, as the results of Experiment 1 suggest that VPE interpretation is never completely categorical in this way. To reconcile the Identity strategy with this observation, it is possible to propose that participants may sometimes make a choice at random. To capture the possible random noise in the data, and also to give the Identity strategy model the best chance to perform well, we introduced a free parameter, ϵ , that allows for some degree of random behavior on the part of the listener. In terms of our experiments, this might represent trials where the participant failed to pay attention, while in actual use, this might represent cases where the listener forgot what the antecedent was, couldn't hear the antecedent, wasn't paying attention, and so on.

In this model, there is a probability of $(1 - \epsilon)$ that the listener will correctly execute the core component of the model described above. However, there is also a probability of ϵ that they will select an interpretation at random from the set of possible messages. This is represented in the model as the reciprocal of the cardinality of the set of possible meanings, M . In the case of our experiments, there are two possible meanings, so this fraction is $\frac{1}{2}$.

To estimate the value of ϵ for which the Identity model would perform the best, we used the `mle`

function in R’s `stats4` package to find the value of ϵ that maximized the 9-condition joint likelihood of the Identity model generating the Experiment 1 data, given the Experiment 2 proportions as the prior probabilities. In other words, we maximized the product of the 9 likelihoods of the model generating the Experiment 1 data in each of the comic strip context/antecedent combinations. The value of ϵ that maximized the joint likelihood is approximately 0.391, and the model predictions generated with ϵ set to this value are plotted against the Experiment 1 results in Figure 6.

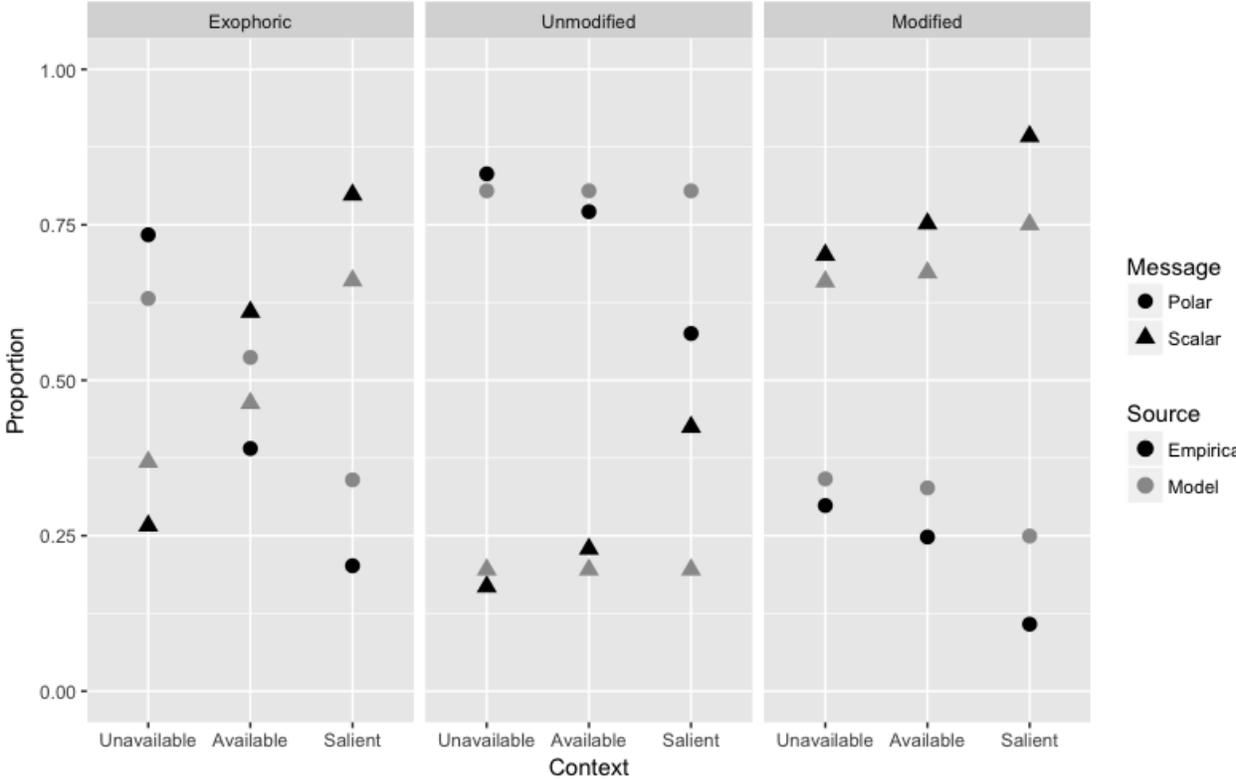


Figure 6: Comparison of observed and predicted VPE interpretation: Experiment 1 vs. Identity model, $\epsilon = 0.391$

In the Exophoric and Modified antecedent conditions, the Identity model predictions show generally the same trends as the empirical observations. This makes sense, as the VPE utterance is marked as compatible with both interpretations. The distribution of interpretations is thus largely driven by the prior discourse status of each interpretation. As shown by the experimental findings earlier, in these conditions, the priors estimated in Experiment 2 largely match listeners’ ultimate judgments for VPE interpretations. Notably, however, even in these conditions where the Identity model performs relatively well, the predicted proportions for the two interpretations were still closer to each other than what was empirically observed. This is because of the high value of ϵ in this instantiation of the model, which will be discussed more below.

The Unmodified antecedent conditions are the most challenging for the Identity model. As mentioned earlier, the noise parameter ϵ is included in Equation 2 to prevent the model from incorrectly ruling out the Scalar interpretation categorically. To maximize the likelihood of the Identity model generating the empirical data, however, ϵ needs to be set to a very high value in order to adjust the model’s otherwise poor predictions in the Unmodified conditions. The predicted proportions shift closer to 0.5 as the value of ϵ is increased. Increasing ϵ from 0 improves the fit of the Identity model in the Unmodified conditions, since the proportions are shifted from the extreme predictions of 1 and 0 relatively closer to 0.5. However, this improvement in fit comes at the expense of fit in the Exophoric and Modified conditions; in these conditions, the prior probabilities are actually closely fit to the posterior interpretation proportions, and the high value of ϵ incorrectly shifts the interpretations closer to 0.5 than is necessary.

Beyond these fit considerations, we note a more practical concern with setting ϵ to a value as high as 0.391, namely that this makes a claim that participants selected an interpretation at random nearly 40% of the time. It seems difficult to justify that listeners choose a random interpretation for an utterance this frequently, either in our task or in naturalistic language comprehension. Finally, even if this were a reasonable value of ϵ , the maximum likelihood instantiation of the Identity model completely fails to predict the Experiment 1 proportions in the Unmodified/Salient condition.

To summarize, a strict Identity strategy fails to provide a non-stipulative account of interpretation in the Exophoric conditions. Even more striking is the finding that the Identity-only strategy also fails to fully account for the conditions in which there is a viable linguistic antecedent (e.g., the Unmodified conditions). Thus, the Identity model formalizes our earlier intuition that a strict Identity account of VPE interpretation provides too small a role for information from the broader discourse context.

5.2. Discourse model

The Discourse model calculates the probability of inferring a particular interpretation upon hearing an utterance using Equation 3. This equation is identical to that used in the Identity model, except that the compatibility ratings are determined using the Discourse strategy rather than the Identity strategy.

$$P(m|u, c) = (1 - \epsilon) \frac{[[u]]_{\text{discourse}}^{m,c} \cdot P(m|c)}{\sum_{m' \in M} [[u]]_{\text{discourse}}^{m',c} \cdot P(m'|c)} + \epsilon \frac{1}{|M|} \quad (3)$$

The Discourse strategy trivially marks the VPE reply as semantically compatible with all alternative target messages. In other words, $\llbracket u \rrbracket_{\text{discourse}}^{m,c}$ is always equal to 1, as is the denominator of the core fraction. The equation therefore can be rewritten as shown in Equation 4.

$$P(m|u, c) = (1 - \epsilon) \cdot P(m|c) + \epsilon \frac{1}{|M|} \quad (4)$$

Thus, according to the Discourse model, the hearer of a VPE utterance has a probability of $(1 - \epsilon)$ of adopting wholesale the prior probability of the meaning under consideration, and a probability of ϵ of choosing an interpretation at random. The joint likelihood of this model is maximized when ϵ is approximately 0.084, and the predicted proportions according to this instantiation of the model are plotted against the empirical posterior proportions in Figure 7.

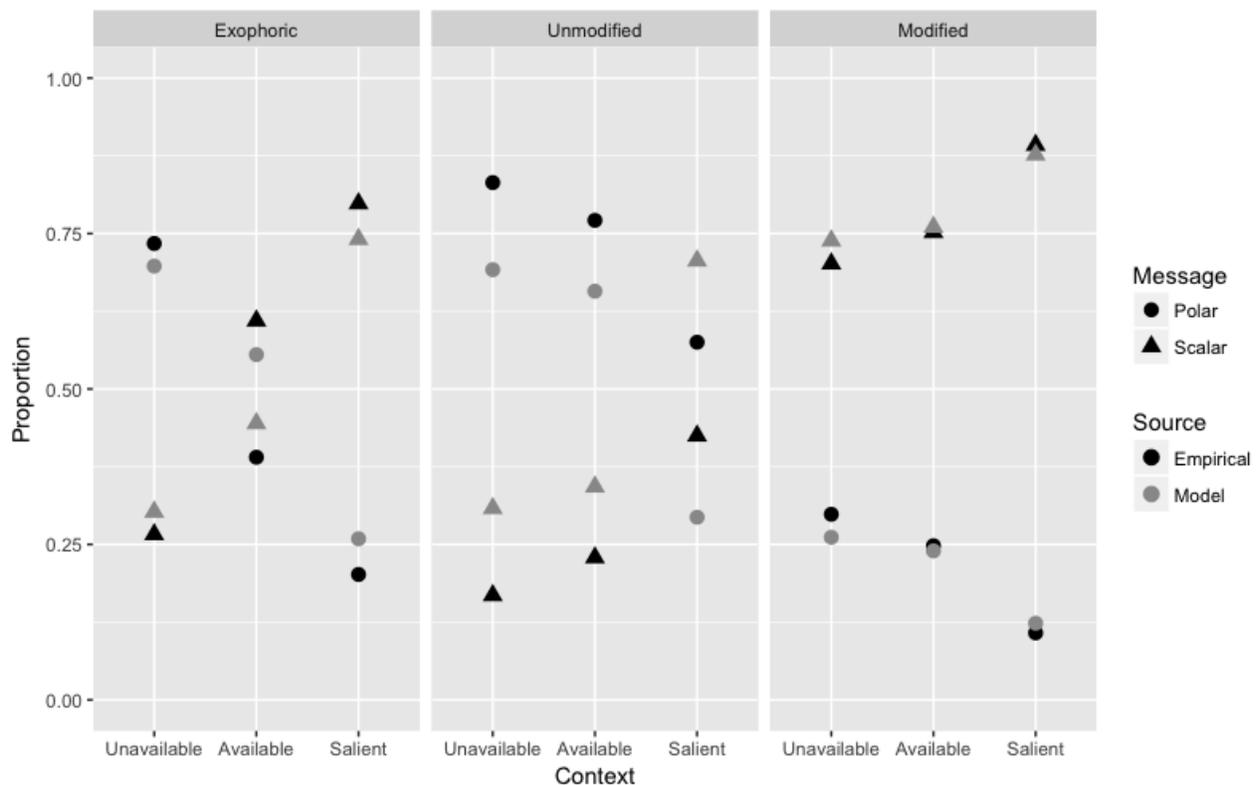


Figure 7: Comparison of observed and predicted VPE interpretation: Experiment 1 vs. Discourse model, $\epsilon = 0.084$

Like the Identity model, the Discourse model makes good predictions for the Exophoric and Modified conditions. In the Unmodified conditions, the Discourse model correctly captures the generalization that making numeral information more available in the comic strip context increases the proportion of times the Scalar interpretation is selected. However, the model systematically underestimates the strength of the Polar interpretation. The mismatch is particularly egregious

with the Unmodified/Salient context, where the model predicts a strong preference for the Scalar interpretation, a poor fit of the empirical observation.

To summarize, since the Discourse model completely discounts the role of linguistic identity, VPE interpretation is predicted to be strictly a function of the prior likelihood of each possible interpretation as determined by general discourse salience. This generates poor empirical predictions; a message that is highly salient in the prior discourse can be penalized if it does not conform to linguistic identity with the antecedent. The Discourse-only model is inherently incapable of capturing this.

5.3. Two-strategy model

The Two-strategy model combines both the Identity and Discourse interpretation strategies. In so doing, it combines the advantages of both strategies, namely by predicting that the distribution of available VPE interpretations is a function of the prior probability of each interpretation, but is also constrained by linguistic identity considerations. There are different ways of instantiating such a model, but the version we will discuss here uses Equation 5 to calculate the probability of inferring a particular interpretation for a VPE utterance.

$$P(m|u, c) = (1 - \epsilon) \left(\beta \cdot \frac{\llbracket u \rrbracket_{\text{identity}}^{m,c} \cdot P(m|c)}{\sum_{m' \in M} \llbracket u \rrbracket_{\text{identity}}^{m',c} \cdot P(m'|c)} + (1 - \beta) \cdot P(m|c) \right) + \epsilon \cdot \frac{1}{|M|} \quad (5)$$

The core part of this model is in the parentheses multiplied by $(1 - \epsilon)$. This core part uses a new parameter, β , to combine the two interpretation strategies when calculating the probability of a particular message. There is a probability of β that the listener will correctly execute the core of the Identity model, and a probability of $(1 - \beta)$ that they will execute the core of the Discourse model, which is simply to use the prior probability of the target message. The model also incorporates random behavior controlled by ϵ , like the previous models. There is a probability of $(1 - \epsilon)$ that the listener will execute the core part of the model described above, and a probability of ϵ that they will choose an interpretation at random.

Put another way, there is a probability of $(1 - \epsilon) \cdot \beta$ that the listener will interpret VPE under identity, a probability of $(1 - \epsilon)(1 - \beta)$ that they will interpret VPE according to the target messages' prior discourse statuses, and a probability of ϵ that they will pick the interpretation at random. The joint likelihood of this model is maximized when β is approximately 0.420 and ϵ is 0.060. The predictions of this model under these conditions are plotted against the empirical observations in Figure 8.

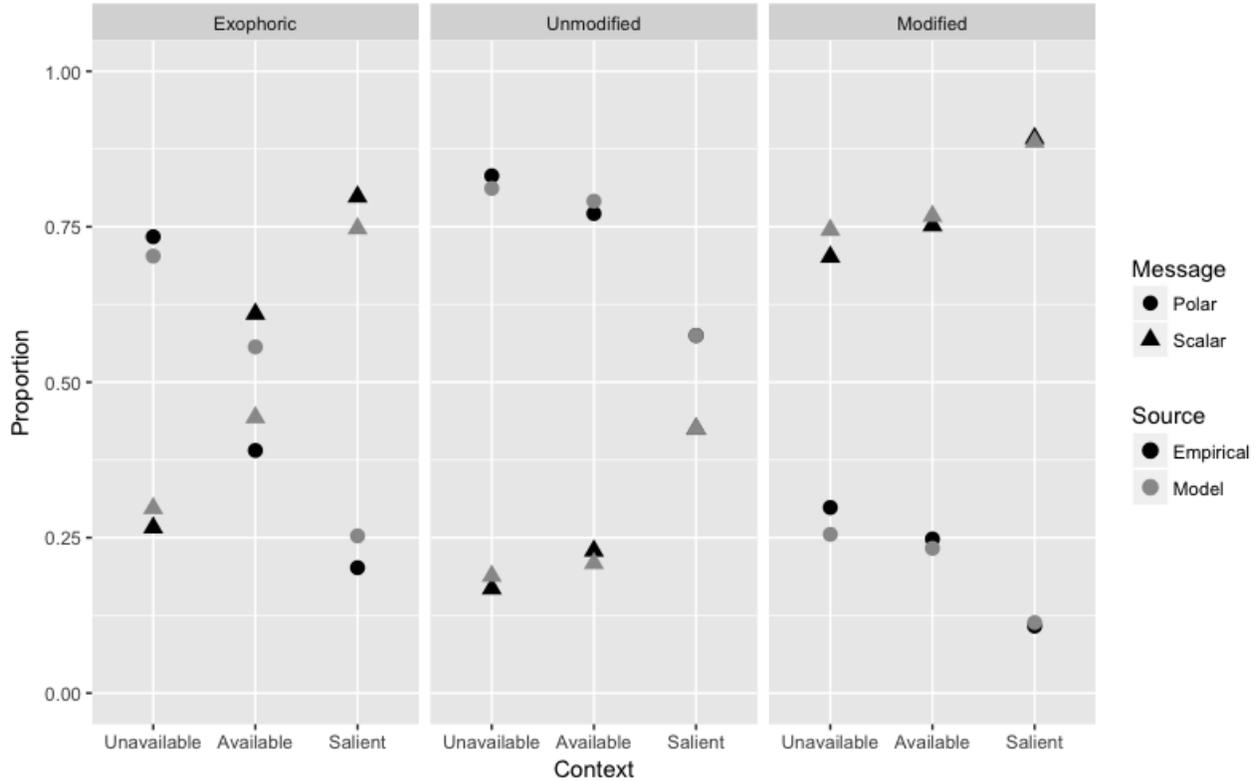


Figure 8: Comparison of observed and predicted VPE interpretation: Experiment 1 vs. Two-strategy model, $\beta = 0.420$, $\epsilon = 0.060$

The Two-strategy model combines the advantages and eliminates the disadvantages of the Identity and Discourse models. While the fit of the model to the empirical data is roughly as good as the other models in the Exophoric and Modified conditions, the fit appears substantially better in the Unmodified conditions. With an Unmodified antecedent, the hybrid model correctly predicts that VPE interpretation is modulated by the salience of the numeral information in the comic strip context, while at the same time the consideration of linguistic identity also correctly constrains the Scalar interpretation when it is highly salient in the discourse context, but incompatible with the antecedent VP.

5.4. Model comparison

Preliminary comparison of the predictions of the three models as described above suggests that the Two-strategy model substantially outperforms both the Identity model and the Discourse model when all conditions are considered. Here, we formalize this conclusion by comparing the models using the Bayes factor (Jeffreys, 1961; Kass and Raftery, 1995; Vandekerckhove et al., 2015). The Bayes factor comparing two models is the ratio of their marginal likelihoods of generating the

observed data. The marginal likelihood of a model generating the data of interest is calculated using Equation 6.

$$\text{ML}_{\mathcal{M}} = \int P(\theta) \cdot P(D|\mathcal{M}, \theta) d\theta \quad (6)$$

In Equation 6, \mathcal{M} represents the model under consideration, D represents the observed data, and θ represents any of the possible combinations of parameter values. Thus, the marginal likelihood for a model is its average likelihood of generating the empirical data, weighted by the prior estimated probability of the combination of parameter values underlying a particular likelihood estimate.

The Bayes factor inherently balances goodness of fit with model parsimony considerations. When a particular model fits the data quite well for many combinations of parameter values, the marginal likelihood is higher, and the model performs relatively better compared to the marginal likelihoods of other models. If a particular parameter only provides a good fit for a very narrow range of possible values, though, the relatively poor fit given the other possible values of the parameter will lower the model’s marginal likelihood, and the model will perform worse when compared to other models. Thus, the Bayes factor should provide a good indication of whether the inclusion of the β parameter in the Two-strategy model compared to the more parsimonious Identity and Discourse models is actually motivated.

Two important considerations when evaluating the marginal likelihood are the distribution of $P(\theta)$ and the bounds of integration - in other words, the probability of each combination of parameter values, and the limits of the parameter values. We will use a flat distribution for $P(\theta)$ because we do not have strong principled expectations about what the exact value of either ϵ or β might be. For the limits of integration, we will consider all values of β from 0 to 1, as we have no prior expectation about what values might be reasonable for this parameter. For ϵ , we will consider values from 0 to 0.2, which is likely a fairly conservative estimate of the maximum proportion of times people select an interpretation at random rather than using the core model calculation to derive the interpretation.¹

As stated above, the Bayes factor is defined as the ratio of the marginal likelihoods of two models, as shown in Equation 7. In the absence of an obvious way to calculate a “joint” marginal

¹We note some concerns regarding the upper bound of ϵ , namely that limiting it to 0.2 excludes the strongest instantiations of the Identity model, the joint likelihood of which is maximized at $\epsilon=0.391$. Increasing the maximum value of ϵ does not change our qualitative conclusion until the maximum value is quite high (e.g., $\epsilon=0.75$, indicating random behavior 75% of the time). With a maximum ϵ of 1, the Bayes factors comparing the Identity and Two-strategy models indicate indifference between the models.

likelihood across all 9 comic strip context/antecedent combinations, we calculated the Bayes factor for each pair of models in each of the 9 conditions separately. The Bayes factor for each pair of models is 1 in all of the Exophoric and Modified conditions, indicating indifference between the models. This is expected, as the models are all mathematically identical in these conditions, with differences in fit determined only by the exact value of ϵ in individual model instantiations. Thus, the Bayes factors for each pair of models in each of the Unmodified antecedent conditions are shown in Table 3. The marginal likelihoods used were approximated using Riemann sums.

$$\text{BF}_{\mathcal{M}_1, \mathcal{M}_2} = \frac{\text{ML}_{\mathcal{M}_1}}{\text{ML}_{\mathcal{M}_2}} \quad (7)$$

Model 1	Model 2	Unavailable	Available	Salient
Two-strategy	Identity	48.0	2.38 e4	2.41 e18
Two-strategy	Discourse	133	12.5	6.26 e6
Discourse	Identity	.361	1.90 e3	3.85 e11

Table 3: Bayes Factors by context for the Unmodified antecedent conditions. A shaded cell denotes that the Bayes Factor is greater than 5, indicating substantial evidence in favor of Model 1.

Cells of Table 3 where the Bayes factor is greater than 5 are shaded. A Bayes factor of 5 indicates that, abstracting over all possible parameter values, Model 1 is 5 times more likely to generate the observed data than Model 2. A Bayes factor of 5 is a conservative threshold for concluding that there is substantially greater evidence in favor of Model 1 than in favor of Model 2. Thus, in all three Unmodified antecedent conditions, the Bayes factors shown in Table 3 indicate that there is substantial evidence that the Two-strategy model captures the Experiment 1 data better than either the Identity or the Discourse model. Since the models make identical predictions in all of the Exophoric and Modified conditions, it is reasonable to conclude from the Unmodified condition Bayes factors that the Two-strategy model is the best model of VPE interpretation in context out of those we have considered.

6. General discussion

The current study interrogated the relative contributions of the linguistic antecedent and the broader discourse context to the interpretation of a VP ellipsis site. The experimental results (Experiment 1) showed that even in the presence of a viable linguistic antecedent, manipulations in the broader discourse context could modulate the distribution of interpretations available for the ellipsis site. Further investigation (Experiment 2) showed that VPE interpretation in context cannot

be reduced to the prior discourse availability of the competing interpretations. To more precisely capture the empirical findings, we constructed and compared three models of VPE interpretation in context. Two of the models were based directly on the Identity and Discourse strategies, augmented with an allowance for random behavior. The third model was a hybrid model that probabilistically combines the Identity and Discourse strategies. The hybrid model provided a substantially better fit for the empirical data than did the pure Identity or Discourse models.

A hybrid account of VPE interpretation in context could help bridge the gap between pure Identity and pure Discourse models of VPE interpretation, representing a first step toward determining the mechanism according to which information from a linguistic antecedent and information from a broader model of the discourse context are combined during the process of verb phrase ellipsis interpretation. This stands in contrast to previous work on VPE interpretation, which has generally focused exclusively on interpretation under strict linguistic identity or interpretation with respect to the discourse context without a separate influence of any linguistic antecedent. The current study expands the empirical domain by examining cases in which both the linguistic antecedent and the broader discourse make available salient information that can potentially be recruited during VPE interpretation. Our finding that information from both sources is indeed recruited during interpretation casts doubt on the “pure” interpretation models and lends strong support to a hybrid account. We note a parallel between this proposal and Frazier and Clifton Jr. (2005)’s argument that both a structure-driven and a discourse-driven processing mechanism are available during ellipsis interpretation. One avenue of future research might focus on developing a unified model of the discourse processor, which Frazier and Clifton recruit to explain non-syntactic salience effects in the linguistic discourse, whereas we rely on the discourse component to explain more abstract salience effects in the discourse context at large.

Ours is not the first account that seeks to model VPE interpretation as a probabilistic process. Bergen and Goodman (2015) show that Rational Speech Act and noisy channel models of utterance interpretation can be combined to model the process of VPE interpretation. Under this account, VPE sentences are inherently ungrammatical in that they constitute an incomplete sentence not generated by the grammar. Speakers choose to use VPE utterances with the knowledge that listeners are equipped with mechanisms to repair utterances that are partly obscured by noise. Listeners, knowing that speakers are aware of these mechanisms, minimally repair ungrammatical VPE utterances by treating the ellipsis site as though it were obscured by noise and making a minimal repair - namely, inferring that the content of the ellipsis site is the previous VP used in the

discourse. We do not view our account as at odds with this model of VPE interpretation. Where Bergen & Goodman’s goal is to provide an account of how VPE sites are interpreted as identical to material available in the prior linguistic context, our account primarily deals with the question of how information from the broader discourse context is incorporated during interpretation. In other words, Bergen & Goodman’s account, as currently construed, is another possible computational implementation of an Identity account, which could in principle be augmented with a component allowing for information from the broader discourse context to contribute to interpretation, and is by no means incompatible with the findings we present here.

The results presented above are relevant not only to the narrow question of verb phrase ellipsis interpretation, but also to the broader theory of verbal anaphora. The traditional view holds that VPE is a “surface” anaphor, in contrast to “deep” verbal anaphors including *do it*, *do this*, *do that*, and null complement anaphora (Hankamer and Sag, 1976; Sag and Hankamer, 1984). On this view, anaphors in the two classes have different structural properties, and the strategies for their resolution are predicted to be categorically different. Previous empirical work has established that formal antecedent-anaphor mismatch leads to greater degradation among surface compared to deep anaphors (Tanenhaus and Carlson, 1990; Mauner et al., 1995). However, more recent work has suggested that the functional difference between these two categories may be gradient rather than categorical (Roberts et al., 2013; Aparicio et al., 2015; Xiang et al., 2019).

The current work provides further evidence for this notion. Contra the observations of Hankamer & Sag, information from outside the linguistic antecedent is not ignored during VPE resolution. Instead, this information plays a role in ellipsis interpretation, albeit one that is constrained by the form of the linguistic antecedent when one is present. This invites a direct comparison between the sensitivity to non-antecedent information during VPE interpretation and the interpretation of traditionally “deep” anaphors. Luce et al. (2018) discuss such a comparison between VPE and *do that* using the same experimental paradigm as the current paper. Their results show that *do that* is relatively more sensitive to information from the broader discourse context than VPE; as numeral information became more salient outside the antecedent, listener’s preferences for the Scalar interpretation of *do that* increased more rapidly than did their preferences for the Scalar interpretation of VPE. Modulo possible differences in situational appropriateness (Miller, 2011), the results suggest that VPE and *do that* represent two positions on a cline of sensitivity to non-antecedent information rather than two opposing sides of a categorical distinction. This observation opens the door for future research on the relative contributions of antecedent and non-antecedent

information to the resolution of various verbal anaphors and the underlying structural and discourse factors that condition their interpretive strategies.

While our results indicate that both linguistic identity and broader discourse information are relevant to VPE interpretation, the exact processing-level mechanism by which information from each source is calculated remains open for further study. Below, we review three possible accounts we view as compatible with our data and model: hybridization in the grammar, parsing ambiguity, and pragmatic accommodation.

Hybridization in the grammar

According to the first account, the grammatical constraint on VPE interpretation takes information from both the linguistic antecedent and the broader discourse context into account, and combines the two information sources in a probabilistic fashion. There are different possibilities for the execution of such a grammatical operation. We discuss two possible instantiations here. The first corresponds to listeners making a stochastic choice between the Identity and Discourse strategies. For any given trial, there is a probability of β that the listener will execute the Identity strategy and draw inference under structural identity with the linguistic antecedent, and a probability of $(1 - \beta)$ that the listener will execute the Discourse strategy. In this scenario, even though both the Identity and the Discourse strategy are available to a listener as grammatical operations, only one of them is executed for a given trial. Which one is chosen is a stochastic decision controlled by the parameter β .² We could imagine that rather than being constant, the exact value of β might be determined by the situation-specific salience of the information available in the nonlinguistic context, such that listeners are more likely to execute the Discourse strategy for interpretation when non-antecedent discourse information is especially salient. This instantiation of the model crucially predicts that at the level of individual trials, listeners make a categorical decision between interpreting an ellipsis site under identity or with respect to the general discourse.

²The instantiation described here is mathematically implemented by the structure of the Two-strategy model from Equation 5, repeated here as Equation 8.

$$P(m|u, c) = (1 - \epsilon) \left(\beta \cdot \frac{\llbracket u \rrbracket_{\text{identity}}^{m,c} \cdot P(m|c)}{\sum_{m' \in M} \llbracket u \rrbracket_{\text{identity}}^{m',c} \cdot P(m'|c)} + (1 - \beta) \cdot P(m|c) \right) + \epsilon \cdot \frac{1}{|M|} \quad (8)$$

In this equation, disregarding the error term, the probability of arriving at a particular interpretation is composed of the product of β and the normalized probability of choosing that interpretation according to the Identity strategy, plus the product of $(1 - \beta)$ and the normalized probability of choosing that interpretation according to the Discourse strategy.

The second possible instantiation is that the Identity and Discourse strategies are both executed and their outputs are fused together for any given trial. This corresponds to a listener who computes the probabilities of an interpretation under both the Identity and the Discourse strategy, and calculates a weighted sum of the two probabilities based on how each strategy is weighted probabilistically. In this scenario, the interpretation of each trial is truly the result of combining information from both sources, rather than the result of a stochastic choice between executing one strategy or the other.³

These two implementations do not make critically different behavioral predictions when examining the aggregated data at the population level. However, as we alluded to above, at the individual trial level, the first implementation predicts a more categorical decision process, where interpretation for an individual sentence is derived either through linguistic identity or through reasoning about the discourse. The second implementation predicts a more gradient outcome, since an individual sentence is interpreted by genuinely combining information from both the linguistic antecedent and the discourse. We will leave it to future work to tease the two apart, but we note there have been some recent methodological proposals on how to model linguistic behavior at the individual trial level (Dillon et al., 2017).

Parsing ambiguity

Even though the findings in this study strongly implicated both the Identity strategy and the Discourse strategy, combining the two information sources in the grammar is only one of the analytical possibilities. The second possibility is to keep only linguistic Identity in the grammar of VPE resolution, but derive the discourse effect by positing a parsing ambiguity regarding the covert structure of the ellipsis site. Merchant (2004, 2016) provides an account of antecedentless VPE according to which listeners infer that the elided content is actually an anaphor like *do that* rather than a contentful VP that can be retrieved from the previous linguistic discourse. In

³To mathematically implement the second possibility, we can modify the equation above into Equation 9:

$$P(m|u, c) = (1 - \epsilon) \frac{\beta(\llbracket u \rrbracket_{\text{identity}}^{m,c} \cdot P(m|c)) + (1 - \beta)(\llbracket u \rrbracket_{\text{discourse}}^{m,c} \cdot P(m|c))}{\sum_{m' \in M} \beta(\llbracket u \rrbracket_{\text{identity}}^{m',c} \cdot P(m'|c)) + (1 - \beta)(\llbracket u \rrbracket_{\text{discourse}}^{m',c} \cdot P(m'|c))} + \epsilon \frac{1}{|M|} \quad (9)$$

In this equation, disregarding the error term, the non-normalized probability of arriving at a particular interpretation is calculated directly as a combination of the probability according to the Identity strategy and the probability according to the Discourse strategy. These two probabilities are combined proportionally according to the value of β . The advantage of this model in terms of Bayes factor compared to the Identity and Discourse models is qualitatively similar to that of the Two-strategy model discussed in the first implementation above.

other words, a string that looks like VPE on the surface could be parsed as either a true VPE construction or a different VP anaphor. The anaphor *do that* (and other similar ones such as *do it*, *do so*, etc.) is known to be very permissive in its ability to be interpreted with respect to the discourse context at large, and so listeners are able to interpret such anaphora as long as there is sufficiently salient information available in the nonlinguistic context. In previous work, we tested overt *do that* anaphora in a similar experimental paradigm to the one reported here, and we indeed found that overt *do that* was interpreted with respect to the general discourse more readily than VPE was (Luce et al., 2018).

The parsing ambiguity account would propose that there is always an ambiguity between parsing an ellipsis site as a silent VP recoverable from a prior linguistic structure and a discourse anaphor like *do that*. There are different possible sources for this parsing ambiguity. The first is that the ambiguity is inherent in the structure of a VPE sentence; that is, the structure of a VPE site is inherently underdetermined, and either parse is available grammatically. Alternately, the VPE site could be, grammatically speaking, unambiguous, but ambiguity could nonetheless arise under parsing frameworks that allow noise or uncertainty in the perceptual input, such as the noisy-channel model of comprehension (Levy, 2008; Gibson et al., 2013; Futrell and Levy, 2017; Ryskin et al., 2018). According to these models, there is always noise or uncertainty regarding the linguistic input a comprehender perceives, and structural alternatives within a reasonable neighborhood of the actual input have a certain probability of being perceived by listeners as the input. A VPE site may inherently cause a relatively large amount of uncertainty regarding the actual input (cf. Bergen and Goodman, 2015), encouraging comprehenders to adopt the “do that” parse instead of the “recoverable VP” parse.

According to the parsing ambiguity account, once the parsing ambiguity arises - either because both parses are inherently available in the input, or because they are close enough perceptual neighbors - the interpretation of the two parses proceeds according to separate mechanisms. The “recoverable VP” parse entails use of a classic Identity strategy mechanism, while the “do that” parse is interpreted with respect to the discourse context at large. The implementation of such an account would by and large look mathematically similar to the Two-strategy model proposed earlier (e.g., Equation 5), except that the Identity component explicitly corresponds to the interpretation strategy used for the “recoverable VP” parse, and the Discourse component corresponds to that used for the “do that” parse. An important question for this approach is to empirically estimate the purported parsing bias, and then fit the parsing ambiguity account to the data observed in the

current study.

Pragmatic accommodation

The last possibility we discuss involves the use of accommodation or repair to arrive at antecedent-nonidentical interpretations. Some accounts of ellipsis interpretation already allow for the inference of a new linguistic antecedent via repair (Arregui et al., 2006) or accommodation (Fox, 1999; van Craenenbroeck, 2013; Thoms, 2015). In these accounts, the construction of a new antecedent is a response to a speaker's use of an ellipsis site that cannot be construed as identical to an overt antecedent; in other words, a formal antecedent-ellipsis mismatch triggers accommodation or repair. Thus, an accommodation account preserves ellipsis interpretation as a process of strict identity, while non-identical interpretations can be derived using a post-grammatical process.

Such accounts can be extended to antecedentless ellipsis in a relatively straightforward way. Antecedentless ellipsis is inherently uninterpretable according to the Identity-driven grammatical interpretation mechanism. Listeners, knowing the speaker intends for the utterance to be interpretable, accommodate or otherwise construct an antecedent using the information available in the broader discourse context, then reinterpret the VPE site with respect to this new antecedent.

Extending such an account to explain the full range of data from the current study is more difficult. In the case of antecedentless ellipsis, there is an obvious trigger for the accommodation or repair operation - VPE, which under the accommodation account is assumed to inherently rely on previous linguistic material for its interpretation, is used in a context where no such material is available, and is thus uninterpretable according to the grammar. In the Unmodified and Modified antecedent conditions of our experiments, by contrast, VPE is used in the context of a completely viable linguistic antecedent. These constructions should be readily interpretable according to the grammatical mechanism, making it unclear why listeners would launch an accommodation or repair operation allowing information from the broader context to intrude on the interpretation of the ellipsis site. This latter pattern is nevertheless exactly what we observed in the critical Unmodified/Salient condition, where salient information from outside the linguistic antecedent was clearly considered during interpretation.

Thus, the accommodation account needs to be substantially modified in order to account for the current data, with a post-grammatical operation allowing non-antecedent information to intrude in VPE interpretation even when there is no clear trigger (e.g., an error signal) for such an operation. Assuming such a theoretical modification were proposed, the accommodation account also makes

testable predictions regarding the different time course under which antecedent versus discourse information starts shaping the interpretation. The accommodation account predicts that listeners will access the antecedent-identical interpretation first, since the grammatical interpretation mechanism for VPE makes reference only to the linguistic antecedent. Only later on in processing, the accommodation operation would activate, allowing for consideration of discourse information.

7. Conclusion

To conclude, the results of the current study provide strong support for a hybrid account of VPE interpretation over a strict Identity-only or Discourse-only model. Our results clearly show that VPE must be examined in rich discourse contexts, with interpretation according to both linguistic antecedents and information from beyond the antecedent considered, in order to arrive at a final, comprehensive description of the interpretive mechanism. The exact realization of the hybrid interpretation mechanism, both in behavioral and model-theoretic terms, is yet to be made more precise by future research, but we charted out a hypothesis space that includes a unified grammatical account, a parsing ambiguity model, and an accommodation or repair mechanism as the potential candidates.

The current study suggests that the interaction between the linguistic system and broader discourse representations is more sophisticated than previously acknowledged in the literature. It is certainly not novel to suggest that language comprehension relies in some way on mental representations of the broader discourse context; certain constructions, like pronouns, are straightforwardly interpreted using salient information from the nonlinguistic context. However, the current study aims to build a more precise quantitative model to account for interpretation in context. The fact that listeners monitor both linguistically and contextually available information during VPE interpretation, and hybridize these information sources in a way that is not reducible to either one in isolation, points to a more entangled connection between the linguistic system and mental models of the context of utterance. Clearly, further research is warranted investigating the boundaries of this connection between linguistic and broader discourse representations.

8. Acknowledgements

Funding: This work was supported by the National Science Foundation [grant number BCS-1827404].

9. References

- Aparicio, H., Franich, K., and Xiang, M. (2015). Morphological feature mismatches under ellipsis: An eye-tracking study. *Proceedings of the 45th Annual Meeting of the North East Linguistics Society*.
- Arregui, A., Clifton Jr., C., Frazier, L., and Moulton, K. (2006). Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language*, 55:232–246.
- Bergen, L. and Goodman, N. D. (2015). The strategic use of noise in pragmatic reasoning. *Topics in Cognitive Science*, 7:336–350.
- Bresnan, J. (1971). A note on the notion 'identity of sense anaphora'. *Linguistic Inquiry*, 2(4):589–597.
- Dalrymple, M., Sheiber, S. M., and Pereira, F. C. N. (1991). Ellipsis and higher-order unification. *Linguistics and Philosophy*, 14:399–452.
- Dillon, B., Staub, A., Levy, J., and Clifton Jr., C. (2017). Which noun phrases is the verb supposed to agree with?: Object agreement in American English. *Language*, 93(1):65–96.
- Fiengo, R. and May, R. (1994). *Indices and Identity*. MIT Press, Cambridge, MA.
- Fox, D. (1999). Focus, parallelism, and accommodation. *Proceedings of the Ninth Conference on Semantics and Linguistic Theory*, pages 70–90.
- Frank, M. C. and Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084):998.
- Franke, M. (2009). *Signal to act: Game theory in pragmatics*. Unpublished doctoral thesis, Universiteit van Amsterdam, Amsterdam, Netherlands.
- Frazier, L. and Clifton Jr., C. (2005). The syntax-discourse divide: Processing ellipsis. *Syntax*, 8(2):121–174.

- Futrell, R. and Levy, R. (2017). Noisy-context surprisal as a human sentence processing cost model. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 688–698.
- Geiger, J. and Xiang, M. (2017). “Context can!”: Contextual accommodation in exophoric and anaphoric verb phrase ellipses. *Proceedings of the 47th Annual Meeting of the North East Linguistic Society*, 2:35–44.
- Gibson, E., Bergen, L., and Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, 110(20):8051–8056.
- Ginzburg, J. and Sag, I. (2000). *Interrogative Investigations*. CSLI Publications, Stanford, CA.
- Goodman, N. D. and Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, 5(1):173–184.
- Hankamer, J. (1978). On the nontransformational derivation of some null VP anaphors. *Linguistic Inquiry*, 9(1):66–74.
- Hankamer, J. and Sag, I. (1976). Deep and surface anaphora. *Linguistic Inquiry*, 7(3):391–428.
- Hardt, D. (1992). VP ellipsis and contextual interpretation. *Proceedings of the International Conference on Computational Linguistics*, pages 303–309.
- Hardt, D. (1993). *Verb phrase ellipsis: Form, meaning, and processing*. Unpublished doctoral thesis, University of Pennsylvania, Philadelphia, PA.
- Jäger, G. (2011). Game-theoretical pragmatics. In van Benthem, J. F. and ter Meulen, A., editors, *Handbook of Logic and Language*, pages 467–491. Elsevier, London, 2nd edition.
- Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, Oxford, 3rd edition.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.
- Kehler, A. (1993). A discourse copying algorithm for ellipsis and anaphora resolution. *Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics*, pages 203–212.

- Kehler, A. (2002). *Coherence in Discourse*. CSLI Publications, Stanford, CA.
- Kertz, L. (2013). Verb phrase ellipsis: The view from information structure. *Language*, 89(3):390–428.
- Kim, C. S., Kobele, G. M., Runner, J. T., and Hale, J. T. (2011). The acceptability cline in VP ellipsis. *Syntax*, 14:318–354.
- Lassiter, D. and Goodman, N. D. (2017). Adjectival vagueness in a Bayesian model of interpretation. *Synthese*, 194(10):3801–3836.
- Levy, R. (2008). A noisy-channel model of rational human sentence comprehension under uncertain input. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 234–243.
- Luce, K. B., Geiger, J., Xiang, M., and Kennedy, C. (2018). Interpretations of VP anaphora through reference to salient events. *Proceedings of the Linguistic Society of America*, 3(38):1–13.
- Maurer, G., Tanenhaus, M. K., and Carlson, G. N. (1995). A note on parallelism effects in processing deep and surface verb-phrase anaphors. *Language and Cognitive Processes*, 10:1–12.
- Merchant, J. (2001). *The Syntax of Silence*. Oxford University Press, Oxford.
- Merchant, J. (2004). Fragments and ellipsis. *Linguistics and Philosophy*, 27:661–738.
- Merchant, J. (2008). An asymmetry in voice mismatches in VP-ellipsis and pseudogapping. *Linguistic Inquiry*, 39(1):169–179.
- Merchant, J. (2016). One size doesn’t fit all: What fragments can tell us about anaphoricity, and why reductionist monolithic theories fail. Presentation, Saarland University, Saarbrücken, Germany.
- Miller, P. (2011). The choice between verbal anaphors in discourse. *Anaphora Processing and Applications: 8th Discourse Anaphora and Anaphor Resolution Colloquium, DAARC 2011*, pages 82–95.
- Miller, P. and Pullum, G. K. (2013). Exophoric VP Ellipsis. In Hofmeister, P. and Norcliffe, E., editors, *The Core and the Periphery: Data-driven Perspectives on Syntax Inspired by Ivan A. Sag*. University of Chicago Press, Chicago, IL.

- Poppels, T. and Kehler, A. (2018). Overcoming the identity crisis: Novel evidence for a referential theory of verb phrase ellipsis. *Proceedings of the 53rd Annual Meeting of the Chicago Linguistic Society*.
- Poppels, T. and Kehler, A. (in press 2019). Reconsidering asymmetries in voice-mismatched verb phrase ellipsis. *Glossa*.
- Roberts, L., Matsuo, A., and Duffield, N. (2013). Processing VP-ellipsis and VP-anaphora with structurally parallel and nonparallel antecedents: An eye-tracking study. *Language and Cognitive Processes*, 28(1-2):29–47.
- Rooth, M. (1992). Ellipsis redundancy and reduction redundancy. *Proceedings of the Stuttgart Ellipsis Workshop*.
- Ryskin, R., Futrell, R., Kiran, S., and Gibson, E. (2018). Comprehenders model the nature of noise in the environment. *Cognition*, 181:141–150.
- Sag, I. (1976). *Deletion and Logical Form*. Unpublished doctoral thesis, Massachusetts Institute of Technology, Cambridge, MA.
- Sag, I. and Hankamer, J. (1984). Toward a theory of anaphoric processing. *Linguistics and Philosophy*, 7:325–345.
- Schachter, P. (1977). Does she or doesn't she? *Linguistic Inquiry*, 8(4):763–767.
- Tanenhaus, M. K. and Carlson, G. N. (1990). Comprehension of deep and surface verbphrase anaphors. *Language and Cognitive Processes*, 5(4):257–280.
- Thoms, G. (2015). Syntactic identity, parallelism and accommodated antecedents. *Lingua*, 166:172–198.
- van Craenenbroeck, J. (2013). Ellipsis, identity, and accommodation. Unpublished manuscript.
- Vandekerckhove, J., Matzke, D., and Wagenmakers, E.-J. (2015). Model comparison and the principle of parsimony. In Busemeyer, J. R., Wang, Z., Townsend, J. T., and Eidels, A., editors, *The Oxford Handbook of Computational and Mathematical Psychology*, pages 300–317. Oxford University Press, Oxford.

Warner, A. (1985). *The structure of English auxiliaries: A phrase structure grammar*. Indiana University Linguistics Club, Bloomington, IN.

Webber, B. L. (1979). *A formal approach to discourse anaphora*. Garland Publishing, New York, NY.

Xiang, M., Grove, J., and Merchant, J. (in press 2019). Structural priming in production through 'silence': An investigation of verb phrase ellipsis and null complement anaphora. *Glossa*.