

Discourse Expectations and the Grammar of Ellipsis

Laura Kertz

*Department of Cognitive, Linguistic, and Psychological Sciences, Brown University,
Providence, RI 02912-1821*

Dec 24, 2013

Abstract

Verb phrase ellipsis is a form of reduced anaphor in English which omits the verb phrase complement to an auxiliary verb. Most theories of ellipsis assume a functional pressure to elide when conditions support it, but there is very little consensus regarding the nature of those conditions. In the present study we demonstrate how contextual manipulations (both syntactic and semantic) can be used to bias expectations for different types of discourse relations and how those discourse-based expectations in turn bias the likelihood of producing an ellipsis. In two follow-up experiments, we showed that the relative ease of processing and relative acceptability for ellipsis versus an unreduced structure is conditioned on the availability of informative cues to discourse relations. These results introduce novel empirical findings regarding rates of ellipsis occurrence in an open-ended production task and present further evidence in support of theoretical proposals which model the influence of argument structure manipulations on ellipsis as an epiphenomenon of constraints on discourse organization and on processing.

Keywords: ellipsis, parallelism, discourse, coherence, anaphora

1. Introduction

Verb phrase ellipsis (VPE), sometime referred to as *post-auxiliary ellipsis* (PAE), is a particularly well-studied anaphor of English, in which the VP complement to an auxiliary verb is omitted, as exemplified in *Alan can't sing, but Marie can*, where the auxiliary 'can' in the second clause has no complement. (All of the data presented here involve verb phrase complements, but an adjective, determiner, or even prepositional phrase predicate is likewise possible, hence the alternative *PAE* label.) The question posed by structures of this sort is obvious: what makes it possible for a speaker to simply omit a constituent that, in a different context, would have to be made explicit? Presumably, the fact that the ellipsis site is governed by an auxiliary head is crucial, as VPE does not turn up in every language (Goldberg 2005). For English-style VPE the presence of the auxiliary verb creates a constrained syntactic environment, triggering the expectation that some predicate is to follow. When that predicate isn't realized, the hearer is cued that she must infer one. The harder question then is what guides the inferencing process that leads her to the intended interpretation. Obviously that process is context dependent, but in what way?

1.1 *Phrase structure (PS) constraints on ellipsis*

One possible answer is that the hearer searches her memory for a predicate which has been recently uttered and which can be 'copied' into the ellipsis site to supply a meaning for the larger structure (Frazier & Clifton 2001). This hypothesis is quite sensible on its face, and it has the added benefit that it suggests a plausible condition for ellipsis to be 'licensed' by the grammar: a VP may be omitted just in case it is identical to one that has been recently uttered. Indeed, the earliest investigations into the grammar of ellipsis adopted precisely this sort of model, elaborated first within a transformational account, and then later within a theory of the interface between different levels of representation (see Schwabe 2003 for an overview). As the deletion under identity account has evolved, it has retained the basic hypothesis that it is an identity condition that makes ellipsis possible, and, crucially, that the constraints governing ellipsis are situated within the phrase-structure component of the grammar.

The trouble with this otherwise appealing account is that it is readily disproved. Since the earliest investigations it has been known that verb phrase ellipsis can be produced under conditions which violate the proposed identity condition (e.g. Sag 1976:75), leading some to propose alternative formulations of the constraint (e.g. Merchant 2001, Fiengo & May 1994, Hardt 1993). It has also been known for some time that ellipses can occur in the absence of a linguistic antecedent, prompting even further re-evaluation (Miller & Pullum 2013; see also Webber 1979, Hankamer 1978, Schachter 1977). Examples of various problematic cases are provided below in Table 2.

Data like these suggest that while context is needed to *evoke* the intended interpretation of an elided VP, identity between that missing VP and some antecedent linguistic expression is not obligatory. That suggestion shifts the theoretical focus away

from constraints on phrase structure and onto interpretability: perhaps ellipsis is possible as long as the intended meaning is recoverable? But this approach, whatever its appeal, is likewise problematic. It is possible to produce an ellipsis which is fully interpretable, but which is nonetheless judged by native speakers to be degraded. An example of this sort is seen in sentence (c) of Table 2. In order to explain the violation observed for a sentence like this one, a simple constraint on recoverability is insufficient.

-
- a.) The aisles at the Lakewood Wal-Mart are surprisingly packed at 11 p.m. “Can we? Can we?” Vanessa tugs at her mother, pointing to a rack of “Lady and the Tramp” DVDs.
b.) This information could have been released by Gorbachev, but he chose not to.
c.) The problem was looked into by Kim just like Lee did.
-

Table 1. Some problematic cases of verb phrase ellipsis. In (a), the missing verb phrase has no linguistic antecedent (COCA, Miller & Pullum 2013). In (b), the missing verb phrase is not identical to the underlined antecedent, but is judged acceptable (NPR, Hardt 1993). In (c), the missing verb phrase is not identical to the underlined antecedent, but is found to be degraded (Frazier & Clifton 2006).

These apparent contradictions have led various researchers to propose that while discourse is crucial to interpretation, ellipsis is nonetheless governed by an identity constraint within the phrase structure component of the grammar (Frazier & Clifton 2005, Winkler 2005, Hardt & Romero 2004, inter alia; see also Williams 1977). As a corollary to this, it has more recently been proposed that for those cases where speakers appear to violate the conditions on licensing with impunity, an adequate performance model will explain those effects as consequences of online processing and/or the mechanisms guiding memory retrieval (Arregui et al. 2006; Grant et al. 2012). This approach, while promising, makes the provocative claim that any instance of ellipsis which violates the identity condition on licensing—no matter how good it might sound to a native speaker—is ungrammatical and properly categorized as a speech error (Frazier 2008).

1.2 *Recent advances in modeling the grammar of ellipsis*

In contrast to proposals which seek to preserve the deletion-under-identity account of ellipsis, some recent proposals have sought to dispense with the identity constraint, or at least to derive it from other more general principles. Those proposals hold that the effects attributed to an identity constraint, in particular the penalty for non-parallelism observed in some cases of ellipsis, are epiphenomenal and can be explained by appealing to interactions between the phrase structure component of the grammar and other facets of the grammar not specifically related to the question of ellipsis licensing. Within this general framework of analysis, two types of proposals have been advanced, each of which broadens a model of the grammar of ellipsis in different ways.

One approach looks to processing to ask how parsing preferences interact with the grammar to shape acceptability judgments. Kim et al. (2011), for example, proposed that a preference for ellipses that conform to an identity condition follows from the interaction

of two parsing heuristics, one which favors structures where arguments are realized in-situ and another which favors elliptical structures which delete larger (as opposed to smaller) chunks of structure. The proposal crucially assumes a model of phrase structure where features of the verb phrase are encoded on functional nodes encapsulating an embedded, underspecified VP. In this sense, the proposal is theory-dependent; however, the account shares assumptions with certain related accounts, notably Fiengo and May's (1994) 'Reconstruction' model.

The parsing heuristics in the Kim et al. model are characterized as expectations and preferences: the parser will expect that for any elliptical clause it encounters, the subject of that clause is an agent of an active voice verb. Next the parser will postpone 'digging' for an underspecified antecedent, embedded below a series of abstract nodes; instead it will preferentially retrieve an antecedent already marked for voice when one is available and use this antecedent to interpret the ellipsis. The net result then is that the parser will assume the ellipsis 'target' is an active voice structure, and in the event that an active voice antecedent is available, the parser will settle on an interpretation fairly quickly. When no active voice antecedent is available, the parser will take longer to reach an interpretation, as it explores alternative parses (eventually digging down to find a suitable underspecified representation of the antecedent). This chain of parsing events is argued to be the source of a general, but crucially violable, preference for parallel structure across ellipsis-antecedent pairs, whereby the extra time needed to construct a representation for a sentence like (c) in Table 1 above is reflected in reduced acceptability. (The question of why a similar penalty is not observed for (b) is not addressed, but presumably follows from independent principles.)

An alternative approach to broadening the grammar of ellipsis emphasizes instead the role of the discourse context, asking how sentence-level structure supports organization at the discourse level. Two distinct, but related, proposals have been offered in this vein. Kehler (2000, 2002) proposed that the preference for parallel structure emerges specifically for ellipses occurring in coherence relations that highlight a comparison or contrast involving two individuals. (Note that example (c) in Table 1, which contrasts *Kim* and *Lee*, meets this condition, and as predicted, shows a penalty for non-parallelism, while example (a) does not.) The proposed mechanism is one in which surface syntactic structure supplies an iconic cue to establishing coherence relations, with arguments filling parallel thematic roles also appearing in parallel syntactic positions when those arguments are contrastive. Kehler proposed that ellipsis is blocked in the case where the argument structure of the overt (unelided) material is insufficient to support the establishment of coherence, i.e., when structure is non-parallel.

Kertz (2013, 2010) offered an alternative explanation, proposing that the preference for parallel argument structure in discourse contexts where a comparison or contrast relation is operative is mediated by information structure. According to that proposal, a general bias (not restricted to ellipsis) favors structures where the two

arguments implicated in the comparison (or contrast) are both realized as subjects/topics of their respective clauses.¹

Each of these lines of inquiry—the parsing approach and the discourse approach to parallelism—shows promise for moving past the log-jam that faces traditional licensing models. At present, however, there is very little overlap between the two approaches. For example, the parsing model elaborated by Kim et al. is blind to discourse (and thus cannot distinguish between minimal pairs which do and do not show a penalty for non-parallelism). The model itself does not preclude effects of the discourse context, but the focus within the proposal maintains an emphasis on phrase structure and engages primarily with the question of whether effects should be attributed to the grammar or to the parser. Meanwhile, the discourse-based proposals have, for their part, offered explanations for why penalties appear to arise for ellipsis in some contexts and not others, but those proposals have been framed as static constraints on representations, with little or no elaboration of how those constraints shape processing in real time.

1.3 *Some methodological concerns*

Another major advance in the study of ellipsis over recent years has been a shift in the types of empirical evidence brought to bear on theory construction. Where much of the early research on ellipsis placed a premium on native speaker intuitions (consistent with practices in the field at the time), research has of late come to rely more and more on data collected from questionnaires administered to speakers sampled from a larger population and from behavioral experiments. This shift has helped in many respects to clarify the nature of the theoretical issues at stake: a case in point involves the various studies undertaken in the 1990s to test an influential proposal from Sag and Hankamer (1984; see also Hankamer & Sag 1976) with respect to the online processing of anaphors. That research was largely carried out by psycholinguists well-versed in linguistic theory. Since that time, researchers in linguistics have adopted many of the tools and methods from psycholinguistics to address theoretical concerns.

There are, however, two considerations with respect to method which, we suggest, have had a limiting effect on theory development. The most prominent of these is that while empirical investigations have gone a long way in describing conditions under which one ellipsis may be judged better than another, or may be processed more easily than another; they have nonetheless skirted the basic question at issue: when is ellipsis preferable to an *unreduced* form? This follows because the bulk of the research on ellipsis, within both linguistics and psycholinguistics, has relied on data comparing ellipsis to ellipsis across contexts or comparing effects of contextual manipulations on ellipsis and other anaphors (see e.g. Roberts et al. 2013 and citations therein). The crucial comparison, one between ellipsis and an unreduced form, is often missing. Without it, we

¹ For empirical results relevant to both of these proposals see Frazier & Clifton (2006) and Kim & Runner (2009).

are left with little to say on what conditions the choice to elide.² To our knowledge there has been only a limited amount of experimental research that makes this comparison, and so far, the results obtained have led researchers to very different conclusions. For example, Kim & Runner (2009) reported a non-parallelism penalty that was specific to verb phrase ellipsis, while Kertz (2013) reported penalties for verb phrase ellipses and their unreduced alternates. (See also Dickey & Bunger 2011 for non-parallelism penalties in sluiced and non-sluiced structures.)

To this basic point about the need for more studies comparing ellipsis and unreduced forms directly, we add a second concern. To date, experimental research on verb phrase ellipsis has relied primarily on acceptability judgments and measures of reading time (though in some cases judgment latencies have also been taken into account). Investigations of this sort depend upon careful contextual manipulation, and then work backward to draw inferences about the grammar. This approach is well-suited to testing theoretically informed hypotheses, but it runs the risk of being overly constrained by the limitations of ‘laboratory speech’, especially when it comes to generating new hypotheses. An obvious gap in the research on elliptical anaphora, we suggest, is the dearth of evidence culled from natural language spontaneously produced. There are a small handful of exceptions to this in the acquisition and aphasia literatures (Kolk 2001, Delis & Slater 1977, Postman et al. 1977), and there is a limited amount of additional work in the linguistics literature drawing on data culled from corpora (e.g. Miller & Pullum 2013, Hardt 1993). Unfortunately, however, the very fact that corpora are comprised of naturally occurring data (precluding, in most cases, the possibility of satisfactory minimal pairs) has led some researchers to discount the evidence as not representative, or worse, irrelevant to theoretical concerns. (For discussion on this point see Frazier & Clifton 2006.)

1.4 *Integrating discourse and processing into a grammar of ellipsis*

In the current work, we set out to redress limitations of past research with respect to theory (we set a goal of integrating discourse and processing into a single model) and method (we collected data using a variety of paradigms, including a language production task). Our investigation was informed by two research questions. First, we asked what conditions the choice to repeat or to elide within a given context. Next, we asked how those factors relate to the apparent parallelism constraint on ellipsis licensing. Our approach was to ask how aspects of the discourse context, including argument structure and semantic cues to discourse relations, affect the process of mapping from a sentence level parse to a richer representation of the discourse—that is, not just arriving at a representation of constituency, but also recognizing who or what is being talked about and how what is being said now relates back to what was said before. We started from the

² One notable exception to this trend is a line of research comparing ellipsis and deaccenting. Seminal work from Rooth (1992) and Tancredi (1992) has suggested that the two may differ minimally in their licensing conditions.

basic idea that whether an ellipsis is supported in a certain context—that is, whether it is likely to be produced (experiment 1), is easy to understand (experiment 2), and sounds acceptable (experiment 3)—depends on how easy it is to make that mapping from the sentence level parse to the discourse level interpretation. From here we generated two hypotheses.

The first of these holds that semantic cues to whether old information is likely to be repeated may be one factor favoring ellipsis over an unreduced form. We hypothesized that discourse connectives (subordinating conjunctions used to introduce a dependent clause) might play an important role. Specifically, we predicted that ellipsis would be more likely to occur and would be judged more acceptable in a clause introduced by ‘just like’ as opposed to one introduced by ‘because’. Our rationale was that ‘just like’ biases an expectation that old information is likely to be re-mentioned. This expectation, we predicted, would in turn favor the use of a reduced form. It seems obvious that an expectation for old information to be re-mentioned is supported by ‘just like’ (to the point that one needn’t bother to test it), but in fact it has not been demonstrated that such an expectation is implicated in the choice to elide. Indeed, while previous work has demonstrated that discourse connectives can influence acceptability of an ellipsis, it has been assumed that this reflects an orthogonal, confounding effect (Frazier and Clifton 2006; see also Kim & Runner 2009). We raise the question here whether connectives (or better, the relationships they denote) might instead be a factor conditioning the preference for ellipsis over an unreduced structure.

Our second hypothesis deals with parallelism. We started with the observation that argument structure matters for establishing particular types of discourse relations, specifically those which feature a contrast between two arguments (as argued both by Kehler and by Kertz, described above). On these grounds we predicted a correlation between argument structure and discourse relation. Using a sentence completion paradigm, we asked whether participants would be more likely to interpret two arguments as being contrastive when they appear in parallel argument positions. We further predicted that parallel argument structure would ease processing and increase acceptability for structures (elided or not) that encode an argument contrast. This latter prediction is consistent with various previous findings in the literature (Frazier et al. 1984, Murphy 1990, Tanenhaus & Carlson 1990, Knoeferle & Crocker 2009) and is likewise consistent with the proposal that parallelism effects for ellipsis follow not from a licensing condition on phrase structure, but from a more general constraint on discourse organization. What’s not clear, however, from prior research, is the precise nature of the discourse constraint, in particular *why* its effect seems to be amplified for ellipsis (at least for some tasks). One possibility is that the parallelism preference for contrastive arguments is a static constraint on discourse, which feeds into a licensing model of ellipsis, blocking it in environments where the constraint is violated (cf. Kehler 2002). An alternative, however, is that argument structure is a cue that is used in real time to map discourse-level relationships, one that is especially important when the sentence-level input is underspecified (Kertz 2013).

Results from our two online studies suggest the latter. We found that parallel argument structure not only biases production patterns but also interpretation. Moreover, we found that the combined effects of cues to discourse relations (from both discourse connectives and from argument structure) modulated the rate of occurrence of ellipsis across different contexts, the type of ellipsis produced across contexts, and the ease of processing for ellipsis. Indeed, when both types of cues conspired to support a discourse relation that encodes an argument contrast, we found that ellipsis induced no added processing costs beyond what was seen for a fully explicit structure. These findings lend empirical support to the theoretical proposal that the parallelism constraint on ellipsis is an epiphenomenon of the combined effects of constraints on discourse organization and on real time processing. This is welcome news, as another finding from our production study is not easily reconciled with the traditional licensing account: roughly 25% of ellipses produced in our experiment violated the parallelism constraint on ellipsis.

We followed up these online studies with an acceptability study, using native speaker judgments to test the effects of our context manipulations on ellipsis and on unreduced structures. (Only structures encoding an argument contrast were tested.) The pattern of results obtained for ellipsis structures was consistent with past results and compatible with predictions of the standard licensing account, with parallel ellipses rated consistently higher than non-parallel ellipses. But the data for the unreduced structures demonstrate that there is more to the story. For those data, higher ratings for parallel structure were likewise observed. Moreover, a preference for ellipsis over an unreduced structure was seen only in the specific case where the ellipsis was parallel *and* was introduced by the connective ‘just like’, revealing a striking correspondence for results across the production, interpretation, and judgment tasks.

2. Effects of discourse context on anaphor production (experiment 1)

Our first experiment elicited spontaneous language production by supplying participants with written sentence fragments which they were asked to complete. We then analyzed their responses to test two hypotheses. First, we predicted that participants would be more likely to supply a completion to the fragment that contained a reduced anaphor in cases where the fragment contained a connective that biased an expectation for the re-mention of old linguistic material. To that end, the fragments supplied to participants contained either ‘just like’, which we predicted would bias for re-mention of old information, or ‘because’ which we predicted would not preclude re-mention of old information, but also, crucially would not bias for it. Example stimuli are seen in (1)-(2).

- (1) The fire fighters endorsed the strike just like / because the police force ...
- (2) The strike was endorsed by the firefighters just like / because the police force ...

Next we varied the argument structure of the fragment, alternating between active and passive voice. We did so to test our second hypothesis, that participants would be more likely to produce a completion that encodes an argument contrast when arguments

filling parallel thematic roles also appeared in parallel positions. For all stimuli, the fragment ended with an animate noun phrase, for example, ‘the police force’ in (1)-(2) above. That noun phrase was always semantically related to another, preceding noun phrase in the fragment, here ‘the firefighters’, which filled the role of agent. (Verbs were always transitive, and the patient role was filled by an inanimate noun phrase.) We predicted that participants would be more likely to produce completions in which the two animate noun phrases are contrastive following a prompt like (1), where both arguments could potentially be realized as subjects in their own clauses. We predicted they would be less likely to do so following a prompt like (2), where the first agent referenced in the fragment was realized as an oblique. (For a complete stimulus set, see the appendix in §9.1.)

In addition to these two hypotheses regarding anaphora and discourse, we asked two questions specific to verb phrase ellipsis. First we asked whether rates of ellipsis in particular would be higher for prompts containing ‘just like’ as compared to ‘because’. We also asked whether productions of ellipsis would show evidence of a preference for parallel argument structure across antecedent/target pairs and/or a constraint blocking production of non-parallel ellipsis.

2.1 Method

Sixteen item sets were constructed as described for (1)-(2) above. Four stimulus lists were auto-generated using the Ibx software package (Drummond 2010). Experimental items were rotated across lists in a Latin square. In each list, sixteen experimental stimuli were interleaved with 40 distractors, randomly ordered. Participants were randomly assigned to lists.

Fifty-two participants were recruited via Amazon’s Mechanical Turk. All fifty-two were self-reported monolingual English speakers. Participants were instructed to complete sentence fragments to produce grammatical, coherent sentences. They were asked to complete each trial quickly, while still ensuring that their responses made sense. Twelve out of 832 responses were judged to be either incoherent or marginally acceptable and were removed prior to analysis.

2.2 Coding and analysis

Participant responses were coded for three features: the type of anaphor (if any) featured in the response, the type of discourse relation obtaining between the prompt and the response, and the argument structure of the response.

Anaphor coding

Responses containing a form of ‘do so’ or the words ‘said so’ were coded as SO ANAPHORS.³ Responses containing a lexical verb with a null clausal complement were coded as NULL COMPLEMENT ANAPHORS. Responses containing a verb phrase ellipsis (auxiliary verb with no complement) were coded as ELLIPSIS. In some cases the matrix verb of the target was elided, in other cases the ellipsis occurred in an embedded clause, and this was recorded. For purposes of analysis, ‘so’ anaphors, null complement anaphors, and ellipses were grouped together as REDUCED ANAPHORS.

Responses which either repeated the lexical verb from the prompt or which included a predicate synonymous with, or antonymous to, the verb in the prompt were grouped together and coded as REPETITION, SYNONYM, or ANTONYM, respectively. Structures of this type we considered to be anaphoric, though not reduced.

Responses were coded as NO VERB when no verb was included in the completion. This type of completion was common (n=32) following prompts with ‘just like’; but did not occur following ‘because’. We note that a classification of this type of structure as anaphoric is not as clear-cut as in other cases. On one view these might be analyzed as elliptical structures and therefore an instance of a reduced anaphor. Alternatively, they might be analyzed as non-elliptical, but nonetheless anaphoric (our own preferred analysis), as they encode relations that are similar to those expressed by REPETITION responses. This issue does not bear directly on any of the hypotheses tested here, and as such we leave the matter unresolved.

Finally, responses which contained a lexical verb, but did not meet any of the criteria described already were coded as NEW PREDICATES. We treated these responses as non-anaphoric. Examples of completions supplied by participants, coded by type, are supplied in Table 2.

Argument structure coding

For each response (excluding instances with no verb) we recorded whether the matrix verb in the completion supplied by the participant was in active voice, passive voice, or formed a copular predicate (a form of *be* followed by a predicate AP, NP, or even PP). For ellipses we also recorded whether the auxiliary governing the ellipsis site selected for an active or a passive voice complement.

³ There are, of course, a variety of differences between ‘do so’ and ‘say so’, which might suggest that the two should not be classed together. Given their low rate of occurrence, however (<2% of the data), we have grouped them together here for convenience.

<i>Example response</i>	<i>Structure type</i>	<i>Anaphoric</i>	<i>Reduced</i>	<i>Argument contrast</i>	<i>Matrix verb of completion</i>	<i>Elided verb</i>
a.) The accounting firm denied the allegation because the legal team <i>advised them to do so</i> .	'so' anaphor	yes	yes	no	active	--
b.) The accounting firm denied the allegation just like the legal team <i>advised</i> .	null complement	yes	yes	no	active	--
c.) The teaching assistants reviewed the material because the faculty instructor <i>asked them to</i> .	ellipsis (embedded)	yes	yes	no	active	active
d.) The freshman senator defended the bill just like the party leader <i>did</i> .	ellipsis (matrix)	yes	yes	yes	active	active
e.) The allegation was denied by the accounting firm just like the legal team <i>denied any wrongdoings</i> .	repetition	yes	no	yes	active	--
f.) The studio was picketed by the screen writers because the stage actors <i>were on strike too</i> .	synonym	yes	no	yes	copula	--
g.) The fire fighters endorsed the strike because the police force <i>had vetoed it</i> .	antonym	yes	no	yes	active	--
h.) The film was abandoned by the lead designer just like the executive producer <i>before him</i> .	no verb	yes	no	yes	--	--
i.) The studio was picketed by the screen writers because the stage actors <i>were not paid fairly</i> .	new predicate	no	no	no	passive	--

Table 2 Example responses coded for completion type. The portion supplied by the participant is italicized.

Coding discourse relations

Responses were further coded for the nature of the discourse relation holding across clauses, specifically whether it encoded an argument contrast. Our criterion for determining whether a response encoded an argument contrast was whether a common property was predicated of distinct individuals across clauses.⁴ For example, in the elliptical sentence in (3), a common property (*defended the bill*) is predicated of *the freshman senator* and of *the party leader*. Responses of this type were coded as contrastive argument relations and were distinguished from other cases in which not just a property, but an entire proposition (including a predicate and its argument), was shared across clauses. In (4), for example, an event in which *the freshman senator defended the bill* is common across both the matrix and the dependent clauses. Note, however, that there is no common predicate holding of the arguments *the freshman senator* and of *the party leader*, which, in this case, are not contrastive.

(3) The freshman senator defended the bill just like the party leader did.

(4) The freshman senator defended the bill because the party leader told him to.

Finally, note that in both (3) and (4) above, a common verb phrase is missing: *defend the bill*, yet the examples differ in their discourse relations. We highlight this to reiterate that the feature of interest to us was not the syntactic structure of the missing material, but instead the discourse relationship obtaining between the two clauses.

2.3 *Results: Rates of anaphora*

Table 3 reports the proportion of response types observed for each condition. As seen in the table, the type of completion supplied by the participant was clearly influenced by the properties of the prompt. Ellipses and null complement anaphors were disproportionately represented in responses to prompts featuring ‘just like’, while new predicates made up a clear majority of the responses to prompts featuring ‘because’. Statistical analysis confirmed a reliable effect of connective for rates of ellipsis ($t=5.15$) and for reduced anaphors generally ($t=10.05$), though notably rates of ‘so’ anaphors were extremely low, and the effect was driven by NCA and ellipsis responses. The choice of connective also reliably affected rates of new predicates ($t=4.08$), though here the effect ran in the opposite direction, with more new predicates following ‘because’ and fewer following ‘just like’. (For details of the statistical analysis see Appendix §8.1).

⁴ At first blush, it might seem counterintuitive that antonymous predicates should meet our criterion of predicating a common property. Our classification requires the assumption that for antonyms, a single underspecified predicate may be inferred which unifies the two. A similar assumption is typically made with respect to the implicatures drawn in certain de-accenting contexts (cf. Schwarzschild 1999).

	<i>because</i>		<i>just like</i>	
	<i>active</i>	<i>passive</i>	<i>active</i>	<i>passive</i>
'so' anaphor	.03	.03	.00	.00
null complement anaphor (NCA)	.05	.02	.25	.43
verb phrase ellipsis (VPE)	.16	.03	.49	.26
repetition, synonym, or antonym	.08	.07	.04	.08
no verb	.00	.00	.22	.22
new predicate	.67	.84	.00	.01

Table 3. Proportion of responses for each completion type, grouped by condition, for experiment 3. Values within each column sum to 1.

The distribution of response types within each condition was also affected by the voice manipulation. Within 'just like' conditions, a higher proportion of VPE were observed following active prompts, and a higher proportion of NCA was seen following passives; suggesting a tradeoff between the two response types. Within 'because' conditions, higher rates of VPE were likewise observed following active prompts, though counts for VPE were lower overall following 'because'. Notably, the trade-off in this case involved not NCA, but new predicates: for 'because' conditions rates of new predicates were lower following active prompts and higher following passives.

For rates of new predicates, a reliable interaction between voice and connective was observed ($t=4.08$). For ellipses, that interaction was reliable at the 2.0 t value cut-off, though a low p value suggests the effects may in fact be additive. For reduced anaphors generally, no interaction was observed, as the tradeoff between NCA and VPE canceled an effect of voice within 'just like' conditions. The trend toward an interaction between voice and connective for rates of reduced anaphors follows from disparities within the 'because' conditions, where rates of ellipses and NCA were both higher following active prompts. Summary results for these statistical analyses are reported in Table 13 (§8.1).

2.4 Results: discourse relations

When we examined the discourse relations obtaining for responses, we observed a strong correlation between response type and whether the noun phrases in the prompt were interpreted contrastively. Indeed for some response types, a perfect correlation was observed. For example, in repetition, synonym, and antonym responses, it was always the case that a common predicate held of the two animate arguments, as these responses, by definition, met our criterion for contrastive argument relations. Likewise responses categorized as no verb completions always encoded an argument contrast. New predicates, by definition, did not encode a contrast. (For examples of responses in each of these types and their coding, see Table 2, above.)

While ellipses were split between some which did and some which did not encode a contrastive argument relation, this correlated with the structure of the ellipsis,

specifically whether it targeted a matrix verb phrase (in which case arguments were contrastive) or an embedded VP (no contrast).⁵ (This was seen, for example, in examples (3) and (4) described in the section on coding discourse relations.) NCA responses patterned with embedded ellipses in not encoding an argument contrast. ‘So’ anaphors also tended to not to encode argument contrasts.

Given the correlation between discourse relation and structure type, it is possible to characterize the complete data set assembled here by first dividing responses into categories of new versus anaphoric, and next subdividing the anaphoric responses into those that do and do not form an argument contrast, and then finally categorizing by the specific structure. By multiplying across the conditional probabilities for each of these categorizations, it is possible to model quite accurately the overall distribution of response types observed in our data set, as demonstrated in Table 4.

		<i>conditional probability</i>		<i>N</i>	<i>N</i>		
				<i>predicted</i>	<i>observed</i>		
Response	.40	new predicate		328	328		
	.60	anaphoric	.51	argument contrast			
				.43	matrix ellipsis	107	107
				.34	no verb	85	85
				.23	repetition, synonym, antonym	57	57
			.49	no argument contrast			
				.62	NCA	149	149
				.33	embedded ellipsis	82	81
				.05	‘so’ anaphor	12	13

Table 4 Conditional probabilities of occurrence of different responses as a function of discourse relation.

With this understanding of the relationship between discourse relation and anaphor type, the next question to ask is how the experimental manipulations influenced discourse relations. Table 6 reports the percentage of responses in each condition that encoded an argument contrast. As reported in the table, contrastive argument relations were seen in roughly half of responses to ‘just like’ prompts, but in only ten percent of responses to

⁵ This is not a necessary outcome, but instead is due largely to the connectives used in this experiment. With an adversative, it is easy to conceive of a matrix ellipsis with no argument contrast, e.g. *Kenny thought about leaving, but he didn’t*, and likewise an embedded ellipsis that does encode an argument contrast, e.g. *Marilyn broke the piano, but everyone thinks Carol did*.

‘because’. Completions with argument contrasts also appeared at a higher rate in response to active voice, as opposed to passive voice, prompts. Statistical analysis confirmed the reliability of each of these effects (connective $t=6.36$; voice $t=5.81$) on rates of argument contrast relations but no interaction between them. Model results for that analysis are reported in Table 14 (§8.1).

	<i>active</i>	<i>passive</i>
because	0.13	0.08
just like	0.57	0.48

Table 5 Percent of responses in each experimental condition which encoded an argument contrast.

2.5 Results: Argument Structure

When we examined the argument structure of the responses supplied by participants, we observed a strong bias for active voice completions, which held across all response types. This can be seen in Table 6, which reports the argument structure of responses as a function of response type. The lowest rate of active voice completions was observed for new predicates, for which roughly a third of responses appeared in the form of a copular construction. Very few passive voice completions were observed, with the highest rate of occurrence in responses that featured new predicates. For remaining response types, the bias for active completions was near absolute.

	<i>active</i>	<i>cop</i>	<i>passive</i>
‘so’ anaphor	1.00		
null complement anaphor	1.00		
ellipsis	.99		.01
repetition, synonym, or antonym	0.95	0.05	
new predicate	0.66	0.28	0.05

Table 6 Proportion of active, passive, and copular structures for each completion type (excluding responses with no verb). Each row sums to 1.

Taking a closer look specifically at responses which contained an instance of verb phrase ellipsis, we found that, overall, ellipsis was more likely to target a matrix verb than an embedded one (107 vs. 81). When data were further subdivided by connective, we found that higher rates of matrix ellipses were observed with ‘just like’ (96 matrix vs. 49 embedded), and higher rates of embedded ellipses were observed with ‘because’ (32 embedded vs. 11 matrix). Next we found that ellipses overwhelmingly targeted an active voice verb. Only eight instances of ellipses targeting a passive voice verb were observed, the majority of these occurring in the passive, ‘just like’ condition (e.g. *The allegation was denied by the accounting firm just like the legal team said it would be.*) These counts, and their distribution across conditions, are shown in Table 7.

		<i>because</i>		<i>just like</i>	
		<i>active</i>	<i>passive</i>	<i>active</i>	<i>passive</i>
		<i>prompt</i>	<i>prompt</i>	<i>prompt</i>	<i>prompt</i>
matrix ellipsis	active target	10		62	33
embedded ellipsis	active target	26	6	33	10
matrix ellipsis	passive target		1		1
embedded ellipsis	passive target				6

Table 7 Counts for ellipses/responses targeting matrix/embedded verbs in active/passive voice (rows) grouped by experimental condition (columns). Shading marks instances where ellipsis target was not voice-matched to the prompt.

Also indicated in Table 7 is whether the ellipsis supplied by the participant was consistent with a parallelism constraint on ellipsis (no shading) or violated it (shaded). Forty-nine out of 188 ellipses produced by participants formed non-parallel structures, the majority of these occurring in the passive, ‘just like’ condition. Thus, participants produced ellipses where ellipsis targets were not structurally matched to their antecedents at a rate of slightly more than 25%. Of these non-parallel ellipsis productions, roughly one third targeted an embedded verb; the remaining targeted a matrix predicate.

2.6 Discussion

The data collected in this experiment show clearly that features of the discourse context may condition not only the rates at which verb phrase ellipsis is produced, but also the type of relation encoded by the ellipsis. We saw no compelling evidence, however, to suggest that speakers are constrained to produce only ellipses which conform to an identity constraint on antecedent ellipsis pairs.

Discourse expectations

Of the two hypotheses tested in this experiment, the first held that rates of reduced anaphor production could be modulated by the discourse connectives present in the prompt. We found indeed that participants were more likely to supply a completion that linked back to old information following ‘just like’, and to do so using a reduced anaphoric structure. Meanwhile they were more likely to supply a completion with a new predicate following ‘because’. On its face this is perhaps not a surprising result, given the meaning of these two particular connectives, but it is important to keep in mind that while ‘just like’ explicitly encodes a comparison relation, it does not specifically *require* a reduced structure for expressing that relation. Old information might instead be repeated or re-mentioned using different lexical items—but that happened at a rate of only 6% following for prompts with ‘just like’. Similarly while ‘because’ encodes a causal relation, it does not preclude an explanation that re-mentions old information (indeed this happened at a rate of just over 22%). Nonetheless, our data show clearly the choice of

connective biases expectations for old versus new information, and, more to the point, that an expectation for old information creates an environment well-suited to reduced anaphora generally and to VPE specifically.

Our second hypothesis predicted a correlation between argument structure and discourse relation, which was likewise confirmed. Production results showed that participants were more likely to supply a completion that encoded an argument contrast when the prompt they were responding to appeared in active voice. That pattern is consistent with proposals from Kertz (2013) and from Kehler (2000) regarding the relationship between sentence-level structuring of arguments and discourse-level meaning. Both have argued that when a discourse relation is structured by comparing two entities with respect to a common property, there will be a preference for those two entities to be expressed in syntactically parallel positions. In conditions where the prompt was in active voice, this afforded the possibility of supplying a completion where both animate referents were evoked from subject position, and as predicted, participants were more likely to treat the arguments as contrastive in that case. Notably, we also observed an independent effect of the connective manipulation, with higher rates of contrastive interpretations following prompts with ‘just like’.

These findings, taken together, suggest a model where discourse-based expectations, which may be derived from distinct cues in the context, bias the likelihood of production for anaphors of different types, verb phrase ellipsis included. Overt semantic cues may generate an expectation for a particular type of relation (a comparison, for example, or an explanation). Argument ordering then supplies a further cue which may fine-tune the nature of the discourse relation. For example, following ‘just like’ there is a general expectation for a comparison relation, and following an active prompt, the likelihood that that comparison focuses on two individuals goes up; following a passive prompt, that likelihood goes down. (See Table 8 for examples.)

a.) The studio was picketed by the screen writers because the stage actors <i>were not paid fairly</i> .	explanation with new predicate
b.) The stock was purchased by the investment banker just like the financial advisor <i>did</i> .	comparison (argument contrast)
c.) The ruling was challenged by the wide receiver just like the assistant coach <i>told them to</i> .	comparison (no argument contrast)

Table 8. Some examples of responses to sentence completion task and their discourse relations.

Contextual constraints on anaphor production

One insight gained from the current results is that while the influence of these contextual manipulations on expectations for particular discourse relations is probabilistic, the relationship between discourse relations and anaphor types was deterministic (or nearly so, at least in this context). Thus while general predictions regarding anaphor use across contexts are imprecise; once the discourse relation is known, our predictions become much more precise. (See Table 4 above.) This finding

echoes recent results from Miller (2011), Miller & Pullum (2013), and Ginzburg (2013), highlighting correspondences between discourse context and anaphor type.

One question left open by the current work, however, is what shapes the apparently contingent relationship between discourse relation and anaphor type seen here. A likely answer seems to involve the choice of predicate supplied in the completion, for which a remarkable regularity was observed. (Object control structure with verbs of saying or thinking, for example, were common for NCA and embedded ellipsis responses.) Whether the likelihood of producing anaphors of different types is modulated by expectations for specific predicates (or classes of predicates) is a promising hypothesis, we suggest, but one which we set aside for future research.

Non-parallel ellipsis production

Before leaving our discussion of these results, we consider briefly how they bear on the question of whether there is an identity constraint governing the licensing of verb phrase ellipsis. The short answer is that the results seen here are not easily reconciled with such an account. There is no evidence, for example, that non-parallel ellipses are categorically blocked; indeed nearly a quarter of the ellipses produced by participants were non-parallel. One might consider that the higher rates of parallel (versus non-parallel) ellipsis productions suggest a violable constraint preferring parallelism, or perhaps a categorical one that is obscured by ‘performance’ related concerns. One might, for example, argue that the non-parallel ellipses are instances of speech errors. Consider, however, that twelve data points were removed from the results prior to analysis because they formed either incoherent or clearly ungrammatical responses. Counts for non-parallel ellipses exceeded all of these bona fide errors combined.

Ultimately it is impossible to assess the claim that a constraint is operative, yet obscured. Instead, we considered the following question: is a parallelism constraint *necessary* to explain the current results. The answer here is no. Indeed, the higher proportion of parallel ellipses seems to follow from two independent trends in the data. First, ellipses tended to target active voice verbs; only 8 instances targeted a passive verb (Table 7). This trend is consistent with the expectation for in-situ argument realization as described by Kim et al. (2011) for ellipsis specifically and for known asymmetries in the realization of voice for English generally (Roland et al. 2007). Next, higher rates of ellipsis were observed following active prompts. (For passive voice prompts, new predicates were the dominant response following ‘because’ and NCA were the dominant response following ‘just like’). From these two trends it follows naturally that higher rates of parallel ellipses were observed.

3. Discourse cues to ellipsis interpretation (experiment 2)

Our second experiment asked how manipulations of argument structure and discourse connectives, which were shown in the previous experiment to affect rates of ellipsis production, might affect online ease of processing. Based on the patterns

observed for the production data, two competing predictions are supported. One places an emphasis on frequency, and predicts that ellipses should be easiest to process in contexts where they are most likely to occur. For the types of data under consideration here (matrix ellipses that invoke an argument contrast) that means that ellipses will be processed most easily when they appear in a clause that is introduced by ‘just like’ and has argument structure that is parallel to the antecedent clause. As a corollary, if it is a frequency-based expectation that facilitates processing, then a tradeoff in ease of processing would be expected for reduced and unreduced structures: where ellipsis is difficult, a structure which repeats instead of eliding should be comparatively easier to process; where ellipsis is relatively easy, repetition should be harder.⁶

But there is a second way of interpreting the production data. Rather than focusing on expectations specifically for ellipsis within a given context, we might ask instead about expectations for different discourse relations and how they support interpretation of ellipsis. This line of inquiry requires a shift away from a view of ellipsis interpretation as a search for an antecedent expression, and toward one of ellipsis interpretation as abductive inference. On the first view, the processing task for interpreting ellipsis is reduced to memory retrieval: accessing a representation of a syn/sem constituent in memory that was recently processed and which can be substituted for the missing structure at the ellipsis site. On the second view, the task at hand involves a mapping from form to meaning: taking an underspecified form as input and mapping it onto a fully specified meaning. From the latter perspective, rather than triggering a search process which would identify an antecedent, the ellipsis site instead triggers an abductive inference regarding the most likely meaning of the structure, given the preceding context (cf. Elbourne 2008; Webber 1978, *inter alia*). A crucial assumption of an inference-based model is that such inferences may be more or less ‘costly’ depending on the types of cues available to support interpretation. (The corollary in a search-based proposal is that some antecedents may be more readily retrieved than others due to the heuristics guiding search (Arregui et al. 2006); but see Martin & McElree (2008) for evidence against a serial search model; see also Phillips & Parker 2013 for discussion.)

A model of anaphor resolution as abductive inference has been proposed at various points to describe pronoun interpretation (Winograd 1972, Hobbs et al. 1993, Kehler et al. 2008, among others), and we suggest here that such an approach can be fruitfully applied to the question of ellipsis processing. The idea that abductive inferences come at a cost is likewise familiar from the literature on lexical pragmatics, where ‘proof costs’ offer a means of formalizing rankings for potential form-meaning mappings and figure within a theory of pragmatic enrichment for utterances that instantiate low-probability pairs (Blutner 1998). The extension we propose here is one in which cues to discourse relations affect the proof cost for interpreting ellipsis. In the preceding experiment, it was

⁶ The latter scenario is suggestive of the ‘repeated name penalty’ described in the literature on personal pronouns (Gordon et al. 1995, Gordon & Scearce 1995, *inter alia*) and discussed specifically with respect to ellipsis by Kim et al. (2011).

demonstrated that participants are more likely to expect a discourse relation that involves contrastive arguments under two conditions: when the connective explicitly cues a comparison and when the argument structure aligns two semantically related arguments in parallel positions. We predict here that those same cues will make it easier to interpret an ellipsis that encodes an argument contrast (than would be the case in the absence of those cues), and, crucially, that those cues will also make interpretation of an *unreduced* structure easier.

The current experiment sought to test these competing hypotheses. We compared reading times for structures which encode an argument contrast, comparing ellipses to unreduced alternates. We manipulated the context preceding the ellipsis (or unreduced VP) to bias expectations for the contrast relation. If it is found that contextual manipulations facilitate ellipsis processing, but impede processing of an unreduced form, then that result would be consistent with a claim that context generates a specific expectation for ellipsis. If, on the other hand it is the case that contextual manipulations are shown to facilitate processing for both ellipsis and unreduced structures, then this would be consistent with a claim that context can bias expectations for specific types of discourse relations, which in turn facilitate processing.

Based on findings elsewhere in the sentence processing literature, there is already some reason to suspect the latter scenario. There is, for example, evidence that ellipses and pro-form verbal anaphors are similarly affected by contextual manipulations (Murphy 1990, Tanenhaus & Carlson 1990; but see Roberts et al. 2013 and Maunder et al. 1995 for qualification). Likewise it has been shown that for coordinate clauses with no anaphora, parallel argument structure can facilitate processing (Frazier et al. 1984, 2000; Knoeferle & Crocker 2009; but see Sturt et al. 2010). Even more compelling is evidence to show that that facilitative effect can be modulated by the choice of discourse connective (Knoeferle 2012). To date, however, there have been no studies specifically comparing effects of voice manipulations on ellipsis and unreduced structures with a goal of assessing the added costs imposed during online processing of VPE.

3.1 *Materials*

Stimulus sentences comprised three clauses: an antecedent and a target, plus a ‘spillover’ region following the target clause. The target clause either repeated the lexical verb from antecedent clause (repetition condition) or contained an auxiliary verb with no complement (ellipsis condition). The target clause was always active voice, but the voice of the antecedent was varied such that argument structure across clauses was either parallel, as in (5) and (7) below, or non-parallel, as in (6) and (8). For all stimuli, a transitive verb in the antecedent clause combined with an animate agent and an inanimate patient. The subject of the target clause was an animate NP that was semantically related to the agent from the antecedent. The connective used to introduce the target clause was varied between item sets: half of the stimulus sets used ‘just like’, as in (5)- (6), while

the other half used ‘because’, as in (7)-(8). For the full list see appendix §**Error! Reference source not found.**

- (5) The local media covered the debate just like the national outlets did,/covered it, but the reporting was better.
- (6) The debate was covered by the local media just like the national outlets did,/covered it, but the reporting was better.
- (7) The family doctor recommended the treatment because the cardiac specialist did,/recommended it, but the prognosis wasn’t good.
- (8) The treatment was recommended by the family doctor because the cardiac specialist did,/recommended it, but the prognosis wasn’t good.

Stimulus lists were auto-generated using the Linger software package (Roland 2003), and lists were rotated across participant sessions in a Latin square. In each session, sixteen experimental stimuli were interleaved with 60 distracters, randomly ordered.

3.2 *Method*

Thirty-six Brown undergraduates received course credit or payment for participation. All participants were self-reported monolingual speakers of English. In a masked, self-paced reading task with a moving window display, participants pressed the space bar to advance through stimuli one word at a time. The display showed dashes in place of characters, with whitespace between words. Participants were thus able to gauge how far along in the sentence they were, as well as the length and number of upcoming words, but were not able to view upcoming words or words already read.

Six practice questions were presented prior to the experimental trials. A yes/no question testing comprehension was asked after each item, and participants received immediate feedback indicating whether they had answered the question correctly. Participants were instructed to read as naturally as possible, making sure that they understood what was read, and to pay attention to feedback on answers to comprehension questions, treating negative feedback as a cue to read more carefully. Experimental sessions took roughly thirty minutes to complete.

3.3 *Results: Comprehension scores*

The overall mean for comprehension scores of experimental items was high (93%), indicating that participants were attending to the content of the material while reading. Means were somewhat lower for ‘because’ conditions as compared to ‘just like’ conditions, as shown in Table 9. Statistical analysis, however, showed no reliable effects

of the experimental manipulations, confirming that the manipulations did not affect participants' ability to ultimately reach an interpretation of the materials.

		<i>because</i>	<i>just like</i>
parallel	ellipsis	90% (2.65)	94% (1.92)
nonparallel	ellipsis	90% (2.62)	94% (2.02)
parallel	no ellipsis	92% (2.36)	95% (1.80)
nonparallel	no ellipsis	93% (2.15)	94% (2.02)

Table 9. Mean comprehension scores (with standard error) by condition for experiment 2.

3.4 Results: Reading times for words in the target clause

Table 10 supplies a schematic of the word-by-word correspondences used in the analysis of reading times. (For complete details of the analysis, see Appendix §8.2)

	<i>target clause subject</i>	<i>target verb phrase</i>	<i>spill-over</i>
ellipsis	the cardiac specialist	did,	but the prognosis
repetition	the cardiac specialist	recommended it,	but the prognosis

Table 10. Schematic of single word regions analyzed in the reading time study.

While our primary interest in reading times concerned the verb phrase of the target clause and, especially, the 'spill-over' region immediately following the target verb phrase, we conducted analyses of all words in the target clause, including words preceding the ellipsis (or repeated verb phrase) site, in order to assess whether the different conditions afforded comparable baselines going into the verb phrase. Table 11 reports reading times for each of the words making up the subject noun phrase and the verb phrase in the target clause.

	<i>The</i>	<i>cardiac</i>	<i>specialist</i>	<i>did</i>	<i>recommended</i>	<i>it</i>
because parallel	351 (11)	338 (11)	360 (10)	367 (19)	363 (12)	366 (15)
because non-parallel	349 (13)	334 (10)	331 (8)	346 (18)	361 (15)	391 (20)
just like parallel	323 (9)	321 (8)	345 (8)	367 (14)	386 (19)	362 (13)
just like non-parallel	315 (9)	340 (12)	342 (8)	394 (22)	372 (15)	419 (22)

Table 11. Reading times (ms) with standard error for words making up the subject noun phrase (the cardiac specialist) and the verb phrase (did/recommended it) of the target clause

While there was some variability in the regions preceding the target verb phrase, the only reliable effect to emerge was seen on the determiner introducing the subject of the target clause. This short-lived effect followed from the connective manipulation, with longer reading times observed when the preceding connective was 'because' ($t = 3.07$). The effect was eliminated by the following word, however. A further disparity was observed in reading times at the head noun for the 'because' conditions, with longer reading times (by roughly 30 ms. on average) in the parallel condition. No such disparity

was observed for the ‘just like’ conditions. Our initial statistical analysis confirmed a reliable interaction between connective and parallelism, as well as a main effect of parallelism, at the head noun. However, further analysis indicated that the reliability of these effects depended on a single outlier in the parallel ‘because’ condition. When this item is removed, the interaction was no longer reliable ($t=1.19$); nor was the main effect ($t=1.15$). Table 11 below, which reports mean reading times, and Table 15 (§8.2), which reports model estimates, show measures and estimates based on the complete data set, including outliers.⁷

For words making up the verb phrase in the target clause, we analyzed ellipsis and repetition conditions separately, finding an effect of the parallelism manipulation in the repetition conditions only. That effect was seen at the pronoun. For ellipsis conditions, some variability was observed for reading times at the auxiliary verb, with especially long reading times in the non-parallel ‘just like’ condition, but statistical analysis showed no reliable effects of the experimental manipulations. Similarly, for repetition conditions, variability was observed at the lexical verb, with somewhat longer reading times for ‘just like’ conditions, but analysis showed no reliable effects. The only reliable effect to emerge in the verb phrase region of the target clause was seen at the pronominal object in repetition conditions, where longer reading times were observed for non-parallel conditions, with especially long reading times in the ‘just like’ non-parallel condition. Statistical analysis confirmed a simple main effect parallelism ($t=2.11$) with no interaction. Results for the statistical analysis are reported in Table 16 (§8.2).

To sum, these results indicate that while a modest effect of the connective manipulations was observed early in the target clause, that effect was not long-lived, and the processing load across conditions appears to have been comparable going into the target verb phrase. In the target verb phrase region itself, no effect of the parallelism manipulation was observed for ellipsis conditions (rather, those effects began on the following word in the spill-over region, described below). For repetition conditions, however, an effect of the parallelism manipulation was seen as early as the pronoun following the repeated verb. This effect emerged as an advantage for parallel structures.

3.5 *Results: Reading times in ‘spill-over’ region following target clause*

We observed effects of the parallelism manipulation across three words in the ‘spill-over’ region following the target clause, our primary region of interest. There we also observed effects of the connective manipulation, including an interaction between

⁷ A follow-up analysis specifically examining the influence of outliers was conducted in which we calculated Studentized residuals for each measurement included in the LME model. Following removal of measures with a residual greater than 3, a new model was computed. That model showed a trend in the same direction as the original model (with elevated reading rates for the head noun in parallel, ‘because’ conditions), but the interaction between parallelism and connective was no longer reliable. Follow-up analyses of this sort were also carried out for reliable effects observed in the spill-over region, all of which were found to be robust.

ellipsis and connective, which emerged at the second word following the clause boundary. Figures 1 and 2 below plot reading times for the first and second words following the clause boundary, respectively.

As seen in Figure 1, at the first word of the spill-over region (the word immediately following ‘did’ in the ellipsis conditions and following ‘it’ in repetition conditions), an advantage for parallel structures was observed, but only in a subset of conditions. For repetitions, which, recall, showed an advantage already at the pronoun following the repeated verb, that advantage persisted through to the following word (Figure 1), and was especially large for repetitions introduced by ‘because’. Meanwhile when we considered the effect of parallelism on ellipsis conditions, we found an advantage for parallel structure only with ‘just like’. (Subset analysis confirmed that the parallelism effect for ‘because’ repetitions is reliable at $p < .05$, marginal for ‘just like’ ellipses at $p = .0569$, and otherwise not reliable.) Overall, the pattern of reading times at the first word of the spillover region shows elevated reading times for ‘because’ conditions (above 375 ms. on average, as seen in the left side of Figure 1), except where parallelism confers an advantage for structures which repeat a lexical verb; meanwhile reading times are overall faster for ‘just like’ conditions (at or below 375 ms., as seen in the right side of Figure 1), except where non-parallelism leads to longer reading times for ellipsis.

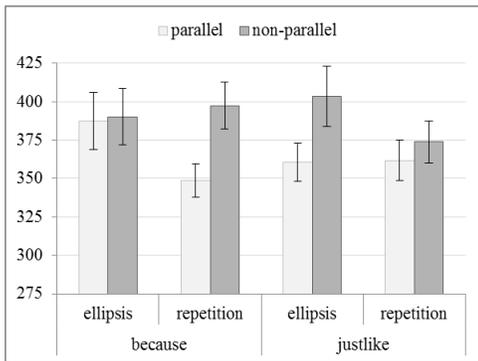


Figure 1 Reading times (ms) by condition at first ‘spillover’ word following clause boundary. Error bars reflect standard error.

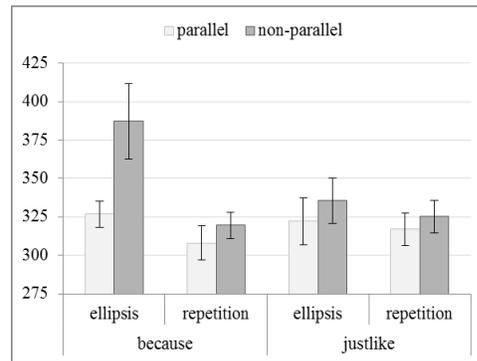


Figure 2 Reading times (ms) by condition at second ‘spillover’ word following clause boundary. Error bars reflect standard error.

By the second word of the spillover region (Figure 2), a small advantage persisted for parallel structures, though reading times dropped off overall, with relatively low reading times observed for seven of the eight conditions. The exception to this trend was seen for non-parallel ellipses introduced by ‘because’ (viz. *The procedure was recommended by the family doctor because the cardiac specialist did, but the prognosis wasn't good.*), for which elevated reading times persisted. This difference is reflected in the statistical analysis as an interaction between connective and ellipsis ($t = 2.56$). (While no 3-way interaction was observed, subset analysis confirmed that an interaction between

parallelism and ellipsis was reliable within the ‘because’ subset only ($p < .01$.) Summary results for analyses of each of these regions are reported in Table 17 (§8.2).

3.6 Discussion

Patterns for reading times showed that the relative ease of processing for ellipsis across conditions were consistent with patterns observed for production: ellipses were easiest to process when the ellipsis followed ‘just like’ and had argument structure parallel to its antecedent, as in (9) below; ellipses were hardest to process when non-parallel and following ‘because’ (10).

- (9) The local media covered the debate
just like the national outlets did / covered it,
but the reporting was better.
- (10) The treatment was recommended by the family doctor
because the cardiac specialist did./recommended it,
but the prognosis wasn’t good.

These results mirrored production patterns where ellipses of this sort (matrix ellipses encoding an argument contrast) occurred most frequently in parallel structures following ‘just like’ and least frequently in non-parallel structures following ‘because’. This consistency confirms that the discourse cues already shown to affect production likewise exert an influence during interpretation, but it does not settle the question of how these cues are exploited. Recall that we have raised two possibilities: either the cues generate a frequency-based expectation for ellipsis, facilitating processing when that expectation is confirmed, or the cues facilitate inferencing about discourse relationships, in turn facilitating ellipsis interpretation. When we compare processing in the ellipsis and non-ellipsis conditions, the results indicate the latter.

To understand this conclusion, consider first the case where ellipsis is the easiest to process: in parallel conditions where the target is introduced by ‘just like’, as in (9) above. If it’s the case, as we propose, that the facilitative effect in this condition follows from converging cues that support a contrastive argument discourse relation, then we would expect to see facilitation for both ellipsis and repetition conditions. And that is indeed what was found. For parallel structures with ‘just like’, reading times at two words following the verb phrase were equivalent whether the structure was elided or not. What this tells us is that the cost for ellipsis in this specific context has been eliminated: when cues are strongly informative with respect to discourse relations, an ellipsis is no harder to interpret than a fully explicit structure. This pattern, while consistent with a model based on discourse expectations, is not the sort of pattern that would be expected if the expectation were instead based on frequency. If cues in the context served to generate a strong expectation for ellipsis *per se*, and not just for a particular type of discourse relation, a penalty would have been observed for the repetition conditions; yet no such penalty was observed.

Now consider the case where ellipsis processing was predicted to be toughest: in non-parallel structures where the elliptical clause was introduced by ‘because’ (10). This is the condition that showed the lowest rates of ellipsis production in experiment 1. Likewise, the reading time results showed clear evidence that these ellipses posed a significant processing challenge. By the second word following the clause boundary, elevated reading times persisted for this condition, in contrast with all other conditions (ellipsis and repetition alike). The question at issue is whether this result indicates that the processing challenge here is specifically one of establishing an interpretation in a context where cues are not informative with respect to discourse relation. If that were the case, we would expect to find a similar difficulty even when the structure is fully explicit. By contrast, if the challenge were simply linked to a frequency-based expectation, we would expect a trade-off with repetition conditions, and ease of processing when the structure was unreduced. No such tradeoff was observed. Evidence that the unreduced version of this structure (the non-parallel repetition introduced by ‘because’) was likewise difficult to process was seen in the first word of the spill-over region. (Compare the bars in Figure 1 for the non-parallel ‘because’ and ellipsis conditions). Even more telling evidence was seen at the pronoun immediately preceding this region in the repetition conditions, where reading times were elevated for non-parallel conditions.⁸

The conclusion we draw from these results is that contextual manipulations do not condition an expectation for ellipsis, specifically, across different contexts, but contextual manipulations *do* condition expectations for different types of discourse relation, and those expectations play a role in supporting the abductive inferences necessary for interpreting ellipsis. This interpretation of the results makes sense of the fact that discourse manipulations affect processing for ellipsis and repetition conditions alike, as in both cases it is necessary to map sentence-level structure onto an integrated discourse representation. The model leaves open the possibility, however, that ellipsis conditions will be disproportionately affected; this follows because in ellipsis conditions, the ‘proof cost’ for mapping an unreduced input onto this representation will be greater. Such a difference in proof costs emerged specifically in the case of non-parallel structures introduced by ‘because’, where cues to the intended discourse relation were insufficient. Notably, however, the model also leaves open the possibility that ellipsis will be no more costly than an unreduced structure, specifically in cases where cues to discourse relations are sufficiently informative. This was likewise observed: in parallel structures introduced by ‘just like’, ellipses incurred no added costs above those seen for an unreduced structure.

⁸ It is of interest here that repetition of a lexical verb does not confer the same advantage when structure is not parallel. This suggests the observed facilitation is not a simple priming effect (either word priming or structure priming, cf. Sturt et al. 2010) but instead involves mapping from sentence-level structures onto a discourse-level meaning. (See Knoeferle & Crocker 2009 for related discussion and findings.)

4. To repeat or to elide (experiment 3)

Our final experiment used a rating scale task to assess speaker preferences for ellipsis versus an unreduced alternate. We asked whether context manipulations shown to affect production and interpretation would likewise affect perceived acceptability for ellipsis structures. At this point in our investigation, a fairly obvious prediction emerges, namely that ellipsis acceptability will be highest in conditions where the ellipsis is introduced by ‘just like’ and argument structure is parallel. Likewise we would predict lowest acceptability when structure is non-parallel and the ellipsis is introduced by ‘because’. Of particular interest here, however, is how those ratings for ellipsis might compare to ratings for unreduced structures. In the present experiment we asked whether contextual manipulations would affect elided and unreduced structures in comparable ways and whether a clear preference for reduced versus unreduced structure might emerge across any of our condition pairings.

Stimulus sets from experiment 2 were adapted for experiment 3, retaining manipulations of connective (‘just like’/‘because’), parallelism (parallel/non-parallel), and ellipsis (ellipsis/repetition); however, the spill-over region following the target clause was eliminated. Example stimuli are provided in (11)-(14) below.

- (11) The local media covered the debate just like.
the national outlets did. / the national outlets covered it.
- (12) The debate was covered by the local media just like
the national outlets did. / the national outlets covered it.
- (13) The family doctor recommended the treatment because
the cardiac specialist did. / recommended it.
- (14) The treatment was recommended by the family doctor because
the cardiac specialist did. / recommended it.

This design allowed us to replicate previous work reported by Frazier & Clifton (2006 experiments 1 & 2), though our design differed from theirs in the inclusion of the non-ellipsis control conditions. The inclusion of these control conditions allowed us to ask specifically about relative preferences for ellipsis versus repetition structures across contexts, as described above. Beyond this, the inclusion of the control conditions permitted us to address additional, unresolved questions, including the role that discourse connectives might play in conditioning those preferences, and whether parallelism effects observed for ellipsis are also seen with unreduced structures. The former question was not addressed in Frazier & Clifton (2006). For the latter question, conflicting results have been reported (Kim et al. 2011, Kertz 2013), as described in the introduction.

4.1 Materials

Thirty-two stimulus sets were constructed which manipulated voice within set, and connective between sets. Thus in sixteen sets, the target clause was introduced by ‘just like’. In the remaining sixteen sets, the target clause was introduced by ‘because’. Four stimulus lists were auto-generated using the Ixweb software package (Drummond 2010). Experimental items were rotated across lists in a Latin square. Within each list, sixteen experimental stimuli were interleaved with 42 distracters, randomly ordered. Participants were randomly assigned to lists.

4.2 Method

Fifty-six participants were recruited via Amazon’s Mechanical Turk. Twelve of these reported to be bilingual and were removed. Stimuli were presented one at a time, and participants were asked to provide a rating for how natural each item sounded, using a scale from 0 (extremely unnatural) to 10 (extremely natural). Ratings and reaction times were both recorded. Six practice trials were completed prior to presentation of the experimental items.

4.3 Results

Acceptability ratings showed clear effects of both the parallelism and the ellipsis manipulations on overall acceptability: parallel structures received consistently higher ratings compared to non-parallel, and repetitions were generally preferred over ellipsis (though not in all cases, as we detail below). Connectives were also seen to affect acceptability, though not in a consistent manner (no general preference for ‘because’ versus ‘just like’); rather, a preference for structures introduced with ‘just like’ arose specifically in ellipsis conditions. Means and standard error for acceptability ratings in each condition are reported in Table 18.

		<i>because</i>	<i>just like</i>
parallel	ellipsis	7.49 (0.17)	7.85 (0.16)
nonparallel	ellipsis	5.22 (0.2)	5.48 (0.2)
parallel	repetition	7.66 (0.17)	7.25 (0.18)
nonparallel	repetition	6.95 (0.2)	6.68 (0.2)

Table 12. Mean acceptability ratings (with standard error) by condition for experiment 3.

As the table shows, higher scores were recorded for parallel conditions across ellipsis and repetition structures alike ($t=6.18$). While that difference was larger for ellipsis conditions (interaction $t=6.97$), subset analysis confirmed that the effect of parallelism was reliable for both ellipsis and repetition groups. The connective manipulation influenced acceptability differently depending on condition: for ellipsis conditions, higher ratings were observed with ‘just like’, for repetition conditions, higher ratings were observed with ‘because’. These opposing trends canceled out such that no

main effect of connective was observed; however, the interaction between ellipsis and connective was reliable ($t=3.02$). Subset analysis of ellipsis data showed a reliable effect of connective for ellipsis conditions; the difference was not reliable, however, for repetitions. A summary of results from the statistical model is reported in Table 18 of §8.3, together with details of the statistical analysis.

The combined effects of these two interactions (one between ellipsis and parallelism, one between ellipsis and connective) yielded a pattern of acceptability as seen in Figure 3, which plots mean acceptability ratings across conditions, normalized by participant. With this scaling, positive scores reflect ratings that are higher than average, as assigned by each participant, and negative scores reflect ratings that are lower than average. As seen in the figure, the only condition in which ellipsis was preferred over repetition was the specific case where argument structure was parallel and the ellipsis was introduced by ‘just like’. When structure was non-parallel, repetition was preferred, and for parallel structures introduced by ‘because’, no preference is observed.

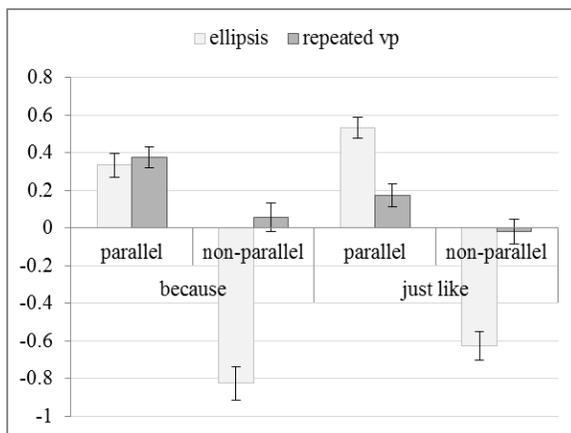


Figure 3 Acceptability ratings for Experiment 3 normalized by participant. Error bars reflect standard error.

4.4 Discussion

The pattern of preferences reflected in speaker judgments of acceptability were consistent with patterns observed for production and interpretation. Ellipses of the sort tested here (contrastive argument ellipses targeting a matrix clause) were judged to be most acceptable in conditions where they were also most likely to be produced (experiment 1) and were easiest to process (experiment 2), that is, when the ellipsis was in a parallel structure and was introduced by ‘just like’, as in (15) below. Conversely, ellipses were judged to be least acceptable in conditions where they were least likely to be produced and where processing demands were greatest, that is, when ellipses appeared in a non-parallel structure introduced by ‘because’ (16).

- (15) The local media covered the debate just like.
the national outlets did. / the national outlets covered it.
- (16) The treatment was recommended by the family doctor because
the cardiac specialist did. / recommended it.

One notable difference emerged, however, between the online measures of processing and the offline acceptability ratings. Where the processing data did not show a tradeoff between ellipsis and repetition conditions, the acceptability data did. This arose specifically in the case where structures were parallel and introduced by ‘just like’ (15). In this condition ellipsis was judged to be the most acceptable, and acceptability for repetition was low by comparison. This pattern contrasted with the reading time pattern, where ellipsis and repetition incurred comparable processing costs. This divergence between offline and online patterns is consistent with past results showing task-dependent effects of contextual manipulations on ellipsis (Murphy 1990, Tanenhaus & Carlson 1990) and highlights the fact that acceptability judgments should not be mistaken for an index of processing cost (a point to which we return in the general discussion).

It furthermore indicates, we propose, that acceptability judgments are sensitive to the availability of competing forms. To illustrate this point we consider just the bars representing parallel conditions in Figure 3. We know from the production data that parallel structures introduced by ‘just like’ offer a prime context for supporting an ellipsis. For structures introduced by ‘because’, however, rates of ellipsis production were relatively low. When it comes to acceptability, participants rated elided and repeated verb structures joined by ‘because’ comparably, reflecting the fact that the context does not bias strongly for either. For structures introduced by ‘just like’, by contrast, a clear preference in acceptability emerged, reflecting the bias to elide in this context. In this context, ratings for ellipsis were higher than expected (taking the ‘because’ conditions as baseline) and ratings for repetition were lower than expected. (Notably, the parallel, ‘just like’ condition was also the condition where lowest rates of completions with a repeated or related lexical item was seen.)

Before closing our discussion of these results, we highlight two additional findings which have a bearing on the interpretation of past empirical studies of ellipsis acceptability. First, the pattern of results here offers an answer to the question left open by past work regarding the nature of connective effects on ellipsis acceptability. Our result indicates that the influence of connectives on acceptability is not an orthogonal concern, as suggested by Frazier and Clifton (2006); rather, the choice of connective used to introduce a subordinate clause can modulate the relative preference for ellipsis versus repetition. Next, our results shed further light on the nature of the parallelism constraint on ellipsis, as they indicate that parallelism effects, while stronger under ellipsis, are not limited to ellipsis contexts. The finding that argument structure parallelism affects acceptability for ellipsis and non-ellipsis structures alike is consistent with an analysis of parallelism effects which attributes the penalty to violation of a constraint on information packaging (Kertz 2013) and with empirical results reported therein. That account does

not offer an immediate explanation for the fact that the non-parallelism penalty is greater for ellipsis structures, but given the broader pattern of results seen here, it seems reasonable to conclude that this follows from the availability of a competing unreduced form and the fact that ellipsis seems to be favored specifically in contexts where expectations for a contrastive argument discourse relation are highest.

Alternatively, the disproportionate effect of parallelism in ellipsis conditions might be explained by appealing to the commonly assumed identity constraint on ellipsis licensing. Indeed, in a study with a similar design (Kim & Runner 2009), an interaction was likewise observed between parallelism and ellipsis, but there no reliable difference was observed outside of ellipsis conditions, and the authors concluded that the non-parallelism penalty is specific to ellipsis.⁹ For the current results, however, an appeal to an ellipsis-specific constraint requires the burdensome assumption that the parallelism penalty seen for ellipsis has two distinct sources: one a discourse-related violation, the other a PS licensing violation. (A more parsimonious account would instead hold that a single constraint is at play, and its effect is modulated depending on whether the structure is reduced.) Perhaps more relevant to the current discussion, however, interpretation of the various other results reported here does not require positing an identity constraint of this sort (indeed the production data argue specifically against one). Overall, our results are most compatible with a theory where the parallelism constraint on ellipsis arises as an epiphenomenon of discourse-based preferences and expectations.

5. General Discussion

Taking these findings as a whole, we are led to three conclusions regarding the nature of the constraints governing verb phrase ellipsis. First, perhaps not surprisingly, we found that anaphor production is context dependent. This was hardly a debatable point prior to the current work, but here we offer evidence for how two specific manipulations matter for conditioning the production of anaphors of different types. Next, we found that context manipulations which can bias for higher rates of ellipsis production also bias expectations for particular types of discourse relations, and that those discourse-based expectations can be manipulated to affect ease of processing. Those expectations, we argued, influence the ‘proof cost’ of mapping an underspecified structure onto a

⁹ One difference between the Kim & Runner study and the present work is that in the former, the voice manipulation affected the target clause (as opposed to the antecedent), as in ‘Kurt blamed Mario for the terrible performance, so Jon was too.’ Ellipses of this sort seem to be especially hard to interpret, given that the auxiliary verb is potentially compatible with an active progressive complement (cf. Kim et al. 2011). Thus it’s possible that this difficulty led to the sizeable penalties they observed for ellipsis conditions, swamping any potential effect in the no-ellipsis cases. Note that Kim & Runner also included a connective manipulation, and while the observed results don’t speak directly to the current proposal (the connectives that were used—‘and’ and ‘so’— seem not to bias an expectation for re-mention), nonetheless a three-way interaction among ellipsis, parallelism, and connective was found, which was attributed to a coherence effect of the sort predicted by Kehler (2002).

particular meaning, in some cases eliminating added cost entirely, such that elliptical structures are no harder to interpret than their unreduced alternates. Finally, we found that native speaker judgments are finely tuned to the likelihood of particular form-meaning mappings, and that judgment patterns previously used to argue for phrase structural constraints on ellipsis licensing might be better understood to reflect sensitivity to factors implicated in that mapping.

5.1 Anaphor production is context dependent

For several decades it has been assumed by many researchers that verb phrase ellipsis is the result of a deletion operation licensed by the phrase structure component of the grammar (but see, e.g. Jacobson 2003, for an opposing view). For almost as long, researchers have grappled with the question of how those constraints might follow from, or be constrained by, discourse context. (See, e.g., Williams 1977 for an early, influential proposal.) The results from the current study indicate that whether speakers choose to elide, or to produce any of a variety of anaphoric expressions, depends on the types of cues available in the context for anticipating discourse relations.

A relatively simple effect of this sort was seen with the manipulation of semantic cues: participants were more likely to produce an anaphoric expression following a connective like ‘just like’, which supplies a strong cue that old information is likely to be repeated. A somewhat less obvious effect was shown to follow from syntactic cues, specifically from the relative positioning of arguments across clauses: when two referents appeared in parallel argument positions, participants were more likely to assign those referents parallel thematic roles and to produce structures in which those referents were implicated in a contrastive discourse relation. The bias to generate a structure encoding a contrast relation in turn biased higher rates of anaphors compatible with that relation, and these two factors together were shown to influence rates of production for elliptical structures which specifically target a matrix clause predicate.

These results offer insight into the little asked question of why different anaphor types might be favored in different contexts (an area which, we suggest, is ripe for more research; see Miller 2011), and they further inform a question that has been the subject of intense debate, namely the nature of the parallelism constraint on ellipsis. The link between argument structure and discourse relations demonstrated here offers new evidence in support of proposals which attribute parallelism effects to general principles of discourse organization (Kertz 2013, 2010; Kehler 2000, 2002), as opposed to a phrase structural constraint applying specifically to ellipsis.

5.2 Context manipulations affect the ‘proof cost’ of ellipsis interpretation

Further evidence for our claim that parallelism effects for ellipsis follow from discourse-based expectations was seen in patterns of online reading times. Where the production data established that discourse connectives and argument structure can both supply cues to discourse relations, the reading time patterns from experiment 2 showed

that when cues of both types converge to bias a strong expectation for an argument contrast relation, processing costs associated specifically with ellipsis can be eliminated. When the two cues are absent, the discourse expectation for an argument contrast relation is weak, and processing for elided and unreduced structures alike is more costly.

The overall pattern of results from that study are consistent with our view of ellipsis interpretation as a form of abductive inference and of context manipulations as factors influencing the ‘proof cost’ for different form-to-meaning mappings. On this view, context manipulations make it harder to map from a sentence-level to a discourse-level meaning (one that integrates across clauses). This in turn hampers ellipsis interpretation since the latter depends on the former. Moreover, the results from the production and interpretation studies showed a striking correspondence, as ellipses were shown to be easiest to process in just those contexts where they were most likely to be produced and hardest to process in those contexts where they were least likely to occur.

5.3 Acceptability ratings are assessments of form-meaning mappings

The foregoing discussion raises an important issue, however, with respect to the role that empirical data may play in informing theory construction. With the recent uptick in interest in grammar models for verb phrase ellipsis has come a wealth of new data regarding acceptability for ellipsis in various contexts. Recent influential proposals (Arregui et al. 2006, Kim et al. 2011) have started from the assumption that acceptability judgments can serve as an index of ease of processing, and that data collected from judgment tasks may serve as a suitable proxy for online processing measures. Our own results suggest that this is not the case and instead show, consistent with past results, that measures obtained from a meta-linguistic judgment task, which engages the participant in reflection, diverge in nontrivial ways from patterns of results based on behavioral measures (cf. Murphy 1990, Tanenhaus & Carlson 1990; see also Cowart 1997, Schutze 1996 for related discussion.)

Rather, what our results indicate is that a more informative correspondence obtains between production patterns and judgments of acceptability. The pattern of results we observed indicate that rather than tracking ease of processing, acceptability judgments reflect a finely tuned sensitivity on the part of speakers to the likelihood of using a particular form within a particular context to express a particular meaning. Our results further suggest that speakers are sensitive as well to competing form-to-meaning mappings, and that their judgments of acceptability may take the availability of these alternatives into account.

6. Conclusions

From a theoretical perspective, the points just discussed raise a variety of implications for our understanding of the nature of the constraints governing ellipsis. In particular they offer further support for the view that a fully adequate model of ellipsis use requires a conception of the grammar that encompasses more than just phrase

structure. We offer evidence, for example, that parsing constraints (like the expectation for in-situ argument realization cited by Kim and colleagues) and constraints on discourse organization (of the sort proposed by Kehler and by Kertz) have a role to play. Indeed, the model we have proposed argues that expectations linked to discourse structure are a determinant factor guiding online processing, as reflected in both reading times and production patterns.

Finally with respect to the basic empirical question which sparked this investigation—what conditions the choice to elide, our results offer several well-defined instances of the ways in which *context matters*. Our results show clearly, for example, that the choice to reduce goes up when discourse biases an expectation for old information to be re-mentioned. Results also show that the specific type of anaphor used depends on the nature of the anticipated relation between old and new information, and, moreover, that expectations for those relations can be reliably manipulated. These biases in turn affect the ease of processing for a reader encountering an ellipsis and for native speakers called on to assess their acceptability.

7. References

- Arregui, Ana, Charles Clifton, Jr., Lyn Frazier, and Keir Moulton. 2006. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language* 55.232–246.
- Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68.255–278. doi:10.1016/j.jml.2012.11.001.
- Blutner, Reinhard. 1998. Lexical pragmatics. *Journal of Semantics* 15.115–162.
- Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgments*. Sage Publications Thousand Oaks, CA.
- Delis, Dean, and Anne Saxon Slater. 1977. Toward a functional theory of reduction transformations. *Cognition* 5.119–132. doi:10.1016/0010-0277(77)90007-5.
- Dickey, Michael Walsh, and Ann C. Burger. 2011. Comprehension of elided structure: Evidence from sluicing. *Language and Cognitive Processes* 26.63–78.
- Drummond, Alex. 2010. *Ibex (Internet Based Experiments)*. <http://github.com/addrummond/ibex>.
- Elbourne, Paul. 2008. Ellipsis sites as definite descriptions. *Linguistic Inquiry* 39.191–220.
- Fiengo, Robert, and Robert May. 1994. *Indices and Identity*. Cambridge, MA: MIT Press.
- Frazier, L., and Charles Clifton, Jr. 2005. The syntax-discourse divide: processing ellipsis. *Syntax* 8.121–174.
- Frazier, Lyn. 2008. Processing ellipsis: A processing solution to the undergeneration problem. *Proceedings of the 26th West Coast Conference on Formal Linguistics*, 21–32. Cascadilla Press.

- Frazier, Lyn, and Charles Clifton, Jr. 2001. Parsing coordinates and ellipsis: Copy alpha. *Syntax* 4.1–22. doi:10.1111/1467-9612.00034.
- 2006. Ellipsis and Discourse Coherence. *Linguistics and Philosophy* 29.315–346.
- Frazier, Lyn, Alan Munn, and Charles Clifton, Jr. 2000. Processing coordinate structures. *Journal of Psycholinguistic Research* 29.343–370.
- Frazier, Lynn, Lori Taft, Tom Roeper, Charles Clifton, Jr, and Kate Ehrlich. 1984. Parallel structure: A source of facilitation in sentence comprehension. *Memory and Cognition* 12.421–430.
- Ginzburg, Jonathan. 2013. Rethinking Deep and Surface: towards a comprehensive model of anaphoric processes in dialogue. Festschrift for Ivan Sag, ed. by Philip Hofmeister.
- Goldberg, Lotus Madelyn. 2005. Verb-stranding VP ellipsis: A cross-linguistic study. UC Santa Cruz.
- Gordon, Peter C, and Davina Chan. 1995. Pronouns, passives, and discourse coherence. *Journal of Memory and Language* 34.216–231.
- Gordon, Peter C, and Kimberly A Scarce. 1995. Pronominalization and discourse coherence, discourse structure and pronoun interpretation. *Memory & Cognition* 23.313–323.
- Grant, Margaret, Charles Clifton Jr, and Lyn Frazier. 2012. The role of non-actuality implicatures in processing elided constituents. *Journal of Memory and Language* 66.326–343.
- Hankamer, Jorge. 1978. On the nontransformational derivation of some null VP anaphors. *Linguistic Inquiry* 9.66–74.
- Hankamer, Jorge, and Ivan Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7.391–426.
- Hardt, Daniel. 1993. Verb phrase ellipsis: Form, meaning, and processing. Philadelphia, PA: University of Pennsylvania ph.d. dissertation.
- Hardt, Daniel, and Maribel Romero. 2004. Ellipsis and the structure of discourse. *Journal of Semantics* 21.375–414.
- Hobbs, Jerry R, Mark E Stickel, Douglas E Appelt, and Paul Martin. 1993. Interpretation as abduction. *Artificial Intelligence* 63.69–142.
- Jacobson, Pauline. 2003. Binding without pronouns (and pronouns without binding). Resource-sensitivity, binding, and anaphora, ed. by Geert-Jan M Kruijff and Richard T Oehrle, 57–96. Dordrecht; Boston; Norwell, MA: Kluwer Academic Publishers ; Sold and distributed in North, Central, and S. America by Kluwer Academic Publishers.
- Kehler, Andrew. 2000. Coherence and the resolution of ellipsis. *Linguistics and Philosophy* 23.533–575.
- 2002. Coherence, reference, and the theory of grammar. Stanford, CA: CSLI.
- Kehler, Andrew, Laura Kertz, Hannah Rohde, and Jeffrey L Elman. 2008. Coherence and coreference revisited. *Journal of Semantics* 25.1–44.

- Kertz, Laura. 2010. Ellipsis reconsidered. San Diego, CA: UC San Diego ph.d. dissertation.
- 2013. Verb phrase ellipsis: The view from information structure. *Language* 89.390–428. doi:10.1353/lan.2013.0051.
- Kim, Christina, Gregory M. Kobele, Jeffrey T. Runner, and John T. Hale. 2011. The acceptability cline in VP ellipsis. *Syntax* 14.318–354.
- Kim, Christina, and Jeffrey T. Runner. 2009. Discourse structure and parallelism in VP ellipsis. *SALT* 19.
- Knoeferle, Pia. 2012. Conjunction meaning can modulate parallelism facilitation: Eyetracking evidence from German clausal coordination. Ms. University of Bielefeld, ms.
- Knoeferle, Pia, and Matthew W. Crocker. 2009. Constituent order and semantic parallelism in online comprehension: Eye-tracking evidence from German. *The Quarterly Journal of Experimental Psychology* 62.2338–2371. doi:10.1080/17470210902790070.
- Kolk, Herman. 2001. Does agrammatic speech constitute a regression to child language? A three-way comparison between agrammatic, child, and normal ellipsis. *Brain and Language* 77.340–350. doi:10.1006/brln.2000.2406.
- Martin, A.E., and B. McElree. 2008. A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language* 58.879–906.
- Maurer, Gail, Michael K. Tanenhaus, and Greg N. Carlson. 1995. A note on parallelism effects in processing deep and surface verb-phrase anaphora. *Language and Cognitive Processes* 10.1–12. doi:10.1080/01690969508407085.
- Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford Studies in Theoretical Linguistics 1. Oxford: Oxford University Press.
- Miller, Philip. 2011. The choice between verbal anaphors in discourse. *Anaphora Processing and Applications*, ed. by Iris Hendrickx, Sobha Lalitha Devi, António Branco, and Ruslan Mitkov, 7099:82–95. *Lecture Notes in Computer Science*. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-25917-3_8.
- Miller, Philip, and Geoffrey K. Pullum. 2013. Exophoric VP Ellipsis. *Festschrift for Ivan Sag*, ed. by Philip Hofmeister, 5–32. Stanford, CA: CSLI.
- Murphy, Gregory L. 1990. Interpretation of verb phrase anaphora: Influences of task and syntactic context. *The Quarterly Journal of Experimental Psychology* 42.675–692.
- Phillips, Colin, and Dan Parker. 2013. The psycholinguistics of ellipsis. *Lingua*.
- Postman, Whitney, Claire Foley, Lynn Santelmann, and Barbara Lust. 1997. Evidence for Strong Continuity: New experimental results from children's acquisition of VP-ellipsis and bound variable structures. *MIT Working Papers in Linguistics* 31.327–344.
- Roberts, Leah, Ayumi Matsuo, and Nigel Duffield. 2013. Processing VP-ellipsis and VP-anaphora with structurally parallel and nonparallel antecedents: An eye-tracking

- study. *Language and Cognitive Processes* 28.29–47.
doi:10.1080/01690965.2012.676190.
- Roland, Douglas, Frederic Dick, and Jeffrey L Elman. 2007. Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language* 57.348–379.
- Rooth, Mats. 1992. Ellipsis redundancy and reduction redundancy. *Proceedings of the Stuttgart Ellipsis Workshop, 29*: Stuttgart: University of Stuttgart.
- Sag, Ivan. 1976. *Deletion and Logical Form*. Cambridge, MA: MIT Press ph.d. dissertation.
- Sag, Ivan, and Jorge Hankamer. 1984. Toward a theory of anaphoric processing. *Linguistics and Philosophy* 7.325–345.
- Schachter, Paul. 1977. Does she or doesn't she? *Linguistic Inquiry* 8.763–767.
- Schutze, Carson T. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. University of Chicago Press.
- Schwabe, Kerstin. 2003. Exploring the interfaces from the perspective of omitted structures. *The interfaces: Deriving and interpreting omitted structures*, ed. by Kerstin Schwabe and Susanne Winkler, 1–28. John Benjamins Publishing.
- Schwarzchild, Roger. 1999. Givenness, AvoidF, and other constraints on the placement of accent. *Natural Language Semantics* 7.141–177.
- Sturt, Patrick, Frank Keller, and Amit Dubey. 2010. Syntactic priming in comprehension: Parallelism effects with and without coordination. *Journal of Memory and Language* 62.333–351.
- Tancredi, Christopher D. 1992. *Deletion, deaccenting, and presupposition*. Cambridge, MA: MIT ph.d. dissertation.
- Tanenhaus, Michael K., and Greg N. Carlson. 1990. Comprehension of deep and surface verb phrase anaphors. *Language and Cognitive Processes* 5.257–280.
- Webber, Bonnie Lynn. 1979. *A formal approach to discourse anaphora*. Outstanding dissertations in linguistics. New York: Garland Pub.
- Williams, Edwin S. 1977. Discourse and logical form. *Linguistic Inquiry* 8.101–139.
- Winkler, Susanne. 2005. Ellipsis and focus in generative grammar. Vol. 81. *Studies in generative grammar*. Walter de Gruyter.
- Winograd, Terry. 1972. Understanding natural language. *Cognitive Psychology* 3.1–191.

8. Appendix 1: Details of statistical analysis

8.1 Analysis of production data (experiment 1)

To analyze rates of responses for different types of structure, a dependent variable was created which scored responses with the structure in question as a 1 and all other responses as a 0. For example, to assess rates of ellipsis production, all responses containing an ellipsis were coded 1, and all other responses were coded 0. Next, an LME model was computed in JMP (SAS Institute Inc. 2010), which modeled scores as a

continuous variable. Fixed effects in all models crossed voice and connective, and a maximal random effect structure (cf. Barr et al. 2013) was used (crossing participant with voice, connective, and their interaction, and likewise crossing item with voice, connective and their interaction). Fixed effects with an absolute t value greater than 2.0 are reported as significant; p values, computed using the Kenward-Roger approximate F test, are reported for significant fixed effects.

<i>Rates of ellipsis</i>	<i>Estimate</i>	<i>Std Error</i>	<i>t Ratio</i>	<i>Prob>t</i>
Intercept	0.24	0.02	9.92	<.0001*
Voice	0.09	0.01	9.03	<.0001*
Connective	-0.14	0.03	-5.15	<.0001*
Voice: connective	-0.02	0.01	-2.07	=.066
<i>Rates of reduced anaphora</i>	<i>Estimate</i>	<i>Std Error</i>	<i>t Ratio</i>	<i>Prob>t</i>
Intercept	0.44	0.03	15.99	<.0001*
Voice	0.05	0.01	4.87	< 0.01*
Connective	-0.27	0.03	-10.05	<.0001*
Voice: connective	0.03	0.02	1.85	=.0886
<i>Rates of new predicates</i>	<i>Estimate</i>	<i>Std Error</i>	<i>t Ratio</i>	<i>Prob>t</i>
Intercept	0.38	0.02	19.03	<.0001*
Voice	-0.04	0.01	-4.47	<.001*
Connective	0.38	0.02	19.1	<.0001*
Voice: connective	-0.04	0.01	-4.08	<.01*

Table 13. Summary results for LME analysis of rates of ellipsis, reduced anaphors (including ellipsis, null complement anaphors, and ‘so’ anaphors), and new predicates.

<i>Rates of argument contrast</i>	<i>Estimate</i>	<i>Std Error</i>	<i>t Ratio</i>	<i>Prob>t</i>
Intercept	0.32	0.04	8.27	<.0001*
Voice	0.04	0.01	5.81	<.01*
Connective	-0.21	0.03	-6.36	<.0001*
Voice: connective	-0.01	0.01	-1.19	0.26

Table 14 LME results for analysis of rates of completions which feature an argument contrast.

8.2 *Analysis of reading data (experiment 2)*

Separate analyses were conducted for comprehension scores and for reading times. All trials were included in the comprehension analysis; only correct comprehension trials were included in the reading time analyses. Statistical analysis for comprehension scores was carried out by computing a LME model using a maximal random effect structure.

For the reading time analysis, measurements below 100 ms. and above 4,500 ms. (n=33) were removed. These cutoffs were chosen based on visual inspection of aggregate data. Means and standard deviations for remaining data, grouped by participant and by condition, were computed, and measurements falling more than 2.5 standard deviations from the mean were replaced with the measurement at 2.5 sd. Together, these trimming and replacement procedures affected 2.6% of the analyzable data. Statistical analysis of reading times was conducted by computing LME models with maximal random effect structure for multiple single-word regions, as indicated in Table 10 (§3.4). Fixed effects with an absolute t value greater than 2.0 are reported as significant; p values, computed using the Kenward-Roger approximate F test, are reported for significant fixed effects.

The choice of fixed effect structure used for each analysis depended on the position of the word within the sentence. For words making up the subject of the target clause, an initial model was computed with three fixed effects (connective, parallelism, ellipsis) fully crossed. The form of the target (ellipsis or repetition) was included in these initial models to test for spurious effects that might be influencing later regions. In each case, when no reliable effect of ellipsis was found, a second model containing just connective and parallelism was computed. Results reported below are from the simpler models.

Given the differences between realization of the verb phrase in ellipsis and repetition conditions (one contains an auxiliary verb at a clause boundary, the other a lexical verb followed by pronominal) we did not undertake analyses directly comparing words in the verb phrase across ellipsis and repetition conditions. Rather, subset analyses were conducted for each word in these regions (auxiliary verb, lexical verb, pronominal object). Fixed effects for those subset analyses crossed parallelism and connective. Analysis of words making up the spill-over region included three fixed effects (connective, parallel, ellipsis) fully crossed.

<i>Determiner</i>	<i>Estimate</i>	<i>Std Error</i>	<i>t Ratio</i>	<i>Prob>t</i>
Intercept	335.94	15.05	22.32	<.0001
Connective	17.01	5.54	3.07	<0.01
Parallelism	2.90	3.35	0.87	
Connective: parallelism	-2.15	3.23	-0.67	
<i>Adjective</i>	<i>Estimate</i>	<i>Std Error</i>	<i>t Ratio</i>	<i>Prob>t</i>
Intercept	334.17	15.80	21.15	<.0001
Connective	3.75	3.66	1.03	
Parallelism	-4.90	3.57	-1.37	
Connective: parallelism	4.04	4.56	0.89	
<i>Noun</i>	<i>Estimate</i>	<i>Std Error</i>	<i>t Ratio</i>	<i>Prob>t</i>
Intercept	344.83	16.28	21.18	<.0001
Connective	2.03	4.44	0.46	
Parallelism	6.81	2.56	2.66	<.05
Connective: parallelism	5.78	2.58	2.25	=.0544

Table 15 Summary results for LME analysis of reading times for words in words making up the subject noun phrase in target clause. (See fn. 7.)

<i>Did</i>	<i>Estimate</i>	<i>Std Error</i>	<i>t Ratio</i>	<i>Prob>t</i>
Intercept	371.12	20.78	17.86	<.0001*
Connective	-11.09	7.05	-1.58	
Parallelism	-6.13	9.09	-0.68	
Connective: parallelism	10.42	7.99	1.28	
<i>Lexical Verb</i>	<i>Estimate</i>	<i>Std Error</i>	<i>t Ratio</i>	<i>Prob>t</i>
Intercept	371.14	19.38	19.19	<.0001
Connective	-11.14	10.73	-0.95	
Parallelism	-6.21	7.14	0.94	
Connective: parallelism	10.26	8.04	-0.6	
<i>Pronoun</i>	<i>Estimate</i>	<i>Std Error</i>	<i>t Ratio</i>	<i>Prob>t</i>
Intercept	385.52	19.57	19.7	<.0001
Connective	-5.82	7.26	-0.8	
Parallelism	-17.74	8.40	-2.11	<0.05
Connective: parallelism	5.84	7.24	0.81	

Table 16 Summary results for LME analysis of reading times for words making up the verb phrase region of the target clause.

<i>Spill 1</i>	<i>Estimate</i>	<i>Std Error</i>	<i>t Ratio</i>	<i>Prob>t</i>
Intercept	377.81	18.61	20.3	<.0001
Connective	5.04	5.91	0.85	
Parallelism	-13.53	5.68	-2.38	<0.05
Connective: parallelism	-0.50	4.36	-0.12	
Ellipsis	8.72	5.29	1.65	
Connective: ellipsis	0.90	4.14	0.22	
Parallelism: ellipsis	0.78	6.18	0.13	
Connective: parallelism: ellipsis	9.79	5.90	1.66	
<i>Spill 2</i>	<i>Estimate</i>	<i>Std Error</i>	<i>t Ratio</i>	<i>Prob>t</i>
Intercept	328.44	13.60	24.14	<.0001
Connective	5.39	4.59	1.18	
Parallelism	-11.29	4.75	-2.37	<0.05
Connective: parallelism	-6.52	5.72	-1.14	
Ellipsis	12.28	5.05	2.43	<0.05
Connective: ellipsis	8.50	3.32	2.56	<0.05
Parallelism: ellipsis	-7.14	5.11	-1.4	
Connective: parallelism: ellipsis	-5.38	5.19	-1.04	
<i>Spill 3</i>	<i>Estimate</i>	<i>Std Error</i>	<i>t Ratio</i>	<i>Prob>t</i>
Intercept	333.75	13.70	24.35	<.0001
Connective	-2.10	4.03	-0.52	
Parallelism	-6.08	3.30	-1.85	=.089
Connective: parallelism	1.94	3.03	0.64	
Ellipsis	-0.98	3.12	-0.32	
Connective: ellipsis	3.32	3.79	0.88	
Parallelism: ellipsis	0.71	4.54	0.16	
Connective: parallelism: ellipsis	-5.10	3.94	-1.3	

Table 17 Summary results for LME analysis of reading times for words in the spillover regions.

8.3 Analysis of acceptability data (experiment 3)

Recognizing that participant engagement can be a concern for experiments conducted online, we analyzed reaction times in order to screen out participants with extremely long or extremely short average trial times. Mean reaction times were calculated for each participant, averaging across both experimental and filler trials. Two participants with average reaction times falling more than 2.5 standard deviations from the group mean were removed, leaving forty-two participants for whom results are reported below.

Ratings data were then analyzed using a linear mixed effect (LME) model computed in JMP (SAS Institute Inc. 2010). The statistical model included each of the fixed effects (PARALLEL, ELLIPSIS, CONNECTIVE) fully crossed and a maximal random effect structure which crossed participant with each within-participant effect (i.e. PARALLELISM, CONNECTIVE, and ELLIPSIS) as well as each within-participant interaction; item was then crossed with each within-item effect (i.e. PARALLELISM and ELLIPSIS) as well as their interaction. Fixed effects with an absolute t value greater than 2.0 are reported as significant; p values, computed using the Kenward-Roger approximate F test, are reported for significant fixed effects. Planned comparisons for simple effects within a single condition were carried out using subset analysis.

<i>Term</i>	<i>Estimate</i>	<i>Std Error</i>	<i>t Ratio</i>	<i>Prob>t</i>
Intercept	6.82	0.24	28.18	<.0001
Connective	0.01	0.08	0.09	
Parallel	-0.74	0.12	-6.18	<.0001
Ellipsis	-0.31	0.07	-4.25	<.001
Connective:Parallel	-0.01	0.07	-0.09	
Connective:Ellipsis	-0.16	0.05	-3.02	<.01
Parallel:Ellipsis	-0.42	0.06	-6.97	<.0001
Connective:Parallel:Ellipsis	0.03	0.05	0.68	

Table 18. Summary results for LME analysis of acceptability ratings.

9. Appendix 2: Experiment Materials

Available on request. Please contact laura_kertz@brown.edu.