

Journal of Semantics, 0, 2014: 1–40
doi:10.1093/jos/ffu006

A Pragmatic Account of Complexity in Definite Antecedent-Contained-Deletion Relative Clauses

EDWARD GIBSON
MIT

PAULINE JACOBSON
Brown University

PETER GRAFF
MIT

KYLE MAHOWALD
MIT

EVELINA FEDORENKO
MIT

STEVEN T. PIANTADOSI
University of Rochester

Abstract

Hackl, Koster-Hale & Varvoutis (2012; HKV) provide data that suggest that in a null context, antecedent-contained-deletion (ACD) relative clause structures modifying a quantified object noun phrase (NP; such as *every doctor*) are easier to process than those modifying a definite object NP (such as *the doctor*). HKV argue that this pattern of results supports a ‘quantifier-raising’ (QR) analysis of both ACD structures and quantified NPs in object position: under the account they advocate, both ACD resolution and quantified NPs in object position require movement of the object NP to a higher syntactic position. The processing advantage for quantified object NPs in ACD is hypothesized to derive from the fact that—at the point where ACD resolution must take place—the quantified NP has already undergone QR whereas this is not the case for definite NPs. Although in other work it is shown that HKV’s reading time analyses are flawed, such that the critical effects are not significant (Gibson *et al.* submitted), the effect in HKV’s acceptability rating is robust. But HKV’s interpretation is problematic. We present five experiments that provide evidence for an alternative, pragmatic, explanation for HKV’s observation. In particular, we argue that the low acceptability of the *the* / ACD condition is largely due to a strong pressure in the null context to use a competing form, by adding *also* or *same*. This pressure does not exist with quantified

NPs either because the competing form is absent (**every same*) or because the addition of *also* actually degrades the sentence. In support of this interpretation, we show that the difference between the *the* / ACD and *every* / ACD conditions (a) persists even when the relative clause contains no ellipsis and thus nothing is forcing QR; (b) disappears when either *also* or *same* is added; and (c) disappears in supportive contexts. Together, these findings show that HKV's QR hypothesis should be rejected in favor of a pragmatic account.

1 INTRODUCTION

The presence of quantificational noun phrases (NPs; or determiner phrases, DPs¹) in object position as in (1) presents a well-known challenge to the view of compositional semantics whereby the lexical meaning of a verb like *read* is assumed to be of type $\langle e, \langle e, t \rangle \rangle$ hence denoting a relation between two individuals:

(1) John read every book.

Because *every book* presumably is of type $\langle \langle e, t \rangle, t \rangle$ (denoting a set of sets, or a generalized quantifier), it is not of the right type to be the object of *read*. Many solutions have been proposed for resolving this mismatch. For example, Montague (1974) posits that the lexical meaning of *read* is actually of type $\langle \langle \langle e, t \rangle, t \rangle, \langle e, t \rangle \rangle$ so that this verb requires a generalized quantifier in object position (and definite NPs and proper names are also of type $\langle \langle e, t \rangle, t \rangle$). Partee & Rooth (1983) and Hendriks (1993) adopt a variant of this hypothesis by which the lexical meaning of *read* is the simple $\langle e, \langle e, t \rangle \rangle$ type but can shift to the more complex Montague meaning. Additional solutions can be found in Kempson *et al.* (2001) and Barker (2002) among others.

However, a more popular solution (which has its roots in a slightly different form in Bach 1968; McCawley 1970; Lakoff 1971) assumes that the input to the semantics is a level of representation distinct from (1), at which *every book* is raised out of the clause by quantifier raising (QR; May 1977). This operation leaves an empty element (a 'trace' of 'movement'), which corresponds to a variable over individuals which is λ -abstracted over. Using the notation from Heim & Kratzer (1998), the relevant representation for (1) is²

(2) every book [8 [Bill read t_8]]

¹ We will use the term NP rather than DP throughout.

² Hackl *et al.* (2012) assume that QR raises the NP to the right; here we use the more standard representation where it is raised to the left. This choice has no bearing on any of the issues here.

where ‘8’ is a label indicating that lambda-abstraction applies to a trace/variable with an index 8. Thus, the interpretation of $[8 [Bill\ read\ t_8]]$ on any assignment g is $\lambda x[Bill\ read\ x]$, which is taken as argument of the generalized quantifier $[[every\ book]]$ (on g).

Since at least as early as Sag (1976), the existence of antecedent contained deletion (ACD) structures (Bouton 1970) as in (3) has been taken as evidence for the QR type of approach:

(3) John read every book which Bill will.

The general assumption is that (under the relevant interpretation) the meaning of the relative clause *which Bill will* can only be compositionally assembled if the ‘missing’ verb phrase (VP) complement of *will* is *read t* (where there is some index on t). Then the semantics puts the whole thing together by having the trace in object position of *read* bound by *which* (or, put differently, the index of the trace is λ -abstracted over to create a function of type $\langle e,t \rangle$ which combines with $[[book]]$). Moreover, because it is generally assumed that the VP *read t* can be deleted only on the basis of identity with other material (the *antecedent*), then in a case like (3), the only way to supply an identical antecedent is to suppose that the object NP is ‘pulled out’ of the main clause also leaving a trace. This allows the matrix VP *read t* to serve as the identical VP here.³ Given that definite objects also license ACD, as in (4), definite NPs can also be ‘pulled out’ by QR:

(4) John will read the book that Bill will.

However, this is not the only analysis of ACD. For example, Cormack (1984), Evans (1988) (for a different but closely related construction), and Jacobson (1992a, 2008) show that—under assumptions from the Categorical Grammar literature (see, e.g. Steedman 1987)—ACD structures can be interpreted in such a way that only a transitive verb meaning (in this case, the meaning of *read*) is understood in the ellipsis site.⁴ Under this view, no trace/variable is needed in object position, so nothing is required to ‘remove’ the object from the main clause VP so as to create this trace/variable. Thus neither the existence of quantified NPs in object position nor the existence of ACD provides evidence for QR over alternative non-movement accounts.

³ In our remarks here and below, we do not mean to suggest that we necessarily endorse a view of ellipsis by which there is ‘deleted’ material. The exact mechanisms by which a ‘missing’ VP in VP Ellipsis is supplied are not relevant here, and we will thus continue to use the standard terminology of ‘deleted’ material for expository convenience.

⁴ This includes complex transitive verb meanings, as in *John was willing to read every book that Mary was*. Here, under the theories discussed above, *willing to read* can compose up to be a complex transitive verb.

In a recent paper, Hackl, Koster-Hale & Varvoutis (2012; henceforth HKV) revisit the implications of ACD for the QR hypothesis, and present a new argument for the QR analysis. Their argument rests on the following premises: (a) QR is required for the case of quantified NPs like *every book* because of the type mismatch; (b) QR is allowed for definite cases like *the book* but is not required; (c) in order to process the ACD case, however, QR is needed for the definite case as well, and (d) the processor takes the minimal route and posits QR only when needed. Consequently, in the initial processing of a definite object, no QR applies in a case like (4). But the processor can't compute a complete representation for (4) under this analysis because there is an ellipsis site. Therefore, it searches for other possible analyses, whereupon QR is attempted. This makes *read t* available as the representation of the matrix VP, and so the ellipsis can be resolved. But for a case like (3)—which contains a quantified object—QR will have already applied, and so *read t* is already available and can be used to resolve the ellipsis. The prediction, then, is that ACD structures modifying a quantified NP should be easier to process than those modifying a definite NP at the point where the processor is attempting to find the antecedent.

Comparing the processing of quantified and definite NPs modified by ACD structures can thus be used to test between the QR account and other accounts. In particular, if there is extra processing load at the ACD site in the definite case, this effect would be unexplained under an account in which (i) quantified NPs in object position are licensed by, e.g. a type-shift on *read*, and (ii) ACD is allowed in some other way (e.g. Cormack 1984; Evans 1988; Jacobson 1992a). Under such an account, there is no connection between the resolution of ACD and the resolution of the apparent type mismatch engendered by the processor encountering a quantified NP. Thus the processing load for ACD (at the ACD site) in the case of NPs with *the* and NPs with *every* should be the same.⁵

HKV present two studies attempting to provide evidence for the QR account, whereby ACD structures with definite NPs engender an extra processing load at the point where the ACD resolution is necessary. The first study examined stimuli as in (5):

- (5) The understaffed general hospital was negotiating with . . .
- a. every doctor that the non-profit medical organization was
 - b. the doctor that the non-profit medical organization was
- . . . in order to arrange for free vaccination clinics.

⁵ Although see Szabolcsi (2013) for discussion of a way that the HKV results can be mimicked without use of QR. We will not discuss Szabolcsi's account here nor compare it with ours, as we are suggesting that the effect is due entirely to pragmatic considerations.

Sentences like (5a) were rated as more acceptable and were read more quickly in a region following the ACD structure. ACD structures in (5) were compared with relative clause structures without ellipsis as in (6):

- (6) The understaffed general hospital was negotiating with . . .
- a. every doctor that the non-profit medical organization funded
 - b. the doctor that the non-profit medical organization funded
- . . . in order to arrange for free vaccination clinics.

HKV reported less of a corresponding difference in complexity between (6a) and (6b), resulting in a significant interaction of clause type (+ACD, -ACD) and object NP modifier (*the*, *every*) in off-line acceptability ratings, and in reading times in a region following the ACD/relative clause verb.

In their second experiment, HKV investigate two ellipsis sites within a sentence, which they refer to as ‘large’ and ‘small’ ellipsis conditions. Their small ellipsis conditions are similar to their Experiment 1 conditions. Because the results for the large ellipsis conditions are actually unexpected under HKV’s account, we postpone consideration of these conditions until the general discussion.

1.1 *Problems with HKV’s interpretation of their results*

In other work, we show that HKV’s reading time analyses were flawed, with the consequence that the critical effects were not significant (Gibson *et al.* submitted). In addition, whereas the QR hypothesis predicts difficulty associated with the *the* condition in (5b) at or immediately following the ellipsis site *was* (the prediction that HKV focus on), it also predicts a corresponding difficulty in the *every* condition earlier in the sentence, when the determiner *every* is first encountered. If anything, there appears to be a trend in the reverse direction in HKV’s own data. In spite of the lack of significance of HKV’s on-line results, their acceptability rating results are replicated here: there is an interaction between quantifier and ellipsis in end-of-sentence acceptability ratings. In the current study, we therefore focus on understanding the source of the acceptability rating effects on HKV’s and related materials.

1.2 *The sameness hypothesis*

The remainder of this article proposes an alternative pragmatic explanation for the acceptability judgments. Note first a weakness in the design of HKV’s experiments: they failed to control for possible plausibility effects or other potential confounds in their materials, because their control conditions for the ellipsis conditions use a different verb than

in the ellipsis conditions. When the same verb is used in the control conditions, we find an advantage for *every* over *the* even without ellipsis, where nothing forces QR with *the*.⁶ HKV's explanation cannot therefore account for this result. The fact that the effect persists without ellipsis thus motivates an alternative explanation of HKV's effect based on pragmatics: the *sameness hypothesis*.⁷

- (7) The sameness hypothesis (a pragmatic explanation of HKV's interaction between determiner (*the*, *every*) and ACD (+, -)): When the target meaning involves performing the same action on both the matrix clause object and the relative clause object (i.e. in the ACD condition or the same verb condition), there is a (pragmatic) pressure to highlight the 'sameness' by using a better form. Such forms are available in the case of *the*: using either *the same* or by the addition of *also*. In the case of *every*, a competing form is either not available (**every same*) or, in the case of *also*, lowers the acceptability (for reasons we discuss briefly in the discussion following Experiment 3). Hence the pressure from competing forms is not there with *every*.⁸

There are two clear advantages of the sameness hypothesis over the QR hypothesis with respect to explaining ACD effects like those presented by HKV. First, the sameness hypothesis predicts an advantage for the *every* condition over the *the* condition in overall acceptability, not just at the ellipsis site. In particular, because there are better ways to express the meaning associated with the *the*-ellipsis condition (e.g. with *the same* instead of *the*, or with addition of *also*), but not with the *every*-ellipsis condition, the *the*-ellipsis condition should be rated as worse overall, as observed. Second, the sameness hypothesis can explain a

⁶ HKV (p. 166) suggest that QR actually might be forced when the same verb is used in the relative clause as in the matrix clause. We consider this idea in detail below, and argue that indeed nothing would force QR for such examples, under HKV's own QR hypothesis.

⁷ We thank Bart Geurts and an anonymous reviewer for help in arriving at this explanation.

⁸ An anonymous reviewer objects that this explanation for the effect requires the computation of an alternative way to express the meaning and then a comparison between them. The reviewer raises the question about just how such a computation and comparison would work. While we will not spell out here how the alternative form (with *the same* or with *also*) is actually computed and compared, we note that there is a lot of independent evidence that this type of computation and comparison does indeed exist, independent of ACD. As we will discuss below in (8), Kaplan (1984) observes a similar phenomenon for cases that don't involve ellipsis: it is difficult to see any way to pin down the oddness of a sentence like (8a) without making reference to the fact that *also* is absent. This kind of an explanation requires reference to a competing form where the absent material is present. Moreover, much of the literature that goes under the rubric of 'Maximize Presupposition' (Heim 1991) makes use of this kind of competition (see, e.g. Amsili et al. 2012). To say that something is ill-formed because it does not include a marker which signals a presupposition requires reference to the competing form that includes the relevant marker. Hence, there is no mechanism assumed by our hypothesis that is not independently motivated.

potential advantage for *every* over *the* in the no-ellipsis conditions, where nothing forces QR with *the*. This explanation works in the same way as in the ellipsis conditions: there are better ways to express the relevant meaning in the *the* condition, but not in the *every* condition. The sameness hypothesis also explains the interaction between the two factors: because ellipsis resolution directs attention to the fact that the meanings of the verb phrases are the same, the pragmatic pressure to highlight this sameness is stronger in the ellipsis conditions than in the no-ellipsis conditions.

The pressure to draw attention to the ‘sameness’ of actions across clauses can be demonstrated independently of the ACD domain. For example, (8a) is odd because there is a pressure to insert *also* into the second clause, as in (8b), or to add *too*, as in (8c) (Kaplan 1984):⁹

- (8) a. Mary read *Crime and Punishment*, and Bill read *Crime and Punishment*.
 b. Mary read *Crime and Punishment*, and Bill also read *Crime and Punishment*.
 c. Mary read *Crime and Punishment*, and Bill read *Crime and Punishment* too.

Critically, ellipsis appears to exaggerate these effects, just as we argue for HKV’s ACD cases: (9a) is highly degraded and one needs *also* or *too* to make this sound coherent, as in (9b) (Kaplan 1984; Amsili *et al.* 2012):

- (9) a. Mary read *Crime and Punishment* and Bill did.
 b. Mary read *Crime and Punishment* and Bill did too.

Furthermore, it appears that the pressure to highlight the ‘sameness’ of the action depends on the extent to which some prior connection is established between the two VPs in the context. (8) and (9) were given above with no prior context and hence no prior connection between the two VP meanings is established, and so *too* is highly preferred. But now consider (8a) in a richer context. Suppose each student in a Russian literature course was required to read one Dostoyevsky novel, and the speaker is commenting on how popular *Crime and Punishment* is this year. In that case (8a) seems to improve. (9a) still seems quite bad, but with rich enough context and appropriate prosody, an ellipsis case can

⁹ Similar points have been noted in the literature that goes under the rubric of ‘Maximize Presupposition’ (Heim 1991; see especially relevant work by Amsili *et al.* 2012). Whether the pressure to use a form like *the same* in the cases of relevance here can be subsumed under that principle is a matter we leave open here.

be more acceptable too; take for example (10) in this context with heavy contrastive focal stress on *Bill*:

- (10) Mary read *Crime and Punishment*; Sue read *Crime and Punishment*.
And Bill did.

Although not entirely parallel to the ACD materials to be discussed here, these examples suggest that there is independent reason to believe that there is a pressure to highlight the ‘sameness’ (perhaps especially so in null contexts), and that this pressure is stronger for the case of ellipsis.

1.3 Outline of the remainder of the article

We present five experiments which provide evidence for the sameness hypothesis over the QR hypothesis as an explanation for HKV’s acceptability rating effects:

- (1) Experiments 1 and 2. HKV failed to control for possible plausibility effects in their materials or for the possible effect from the pressure from competing better forms, as they do not contrast the case of ellipsis with non-ellipsis cases using the same verb. Once one controls for these, we find an advantage for *every* over *the* even without ellipsis, where nothing forces QR with *the*. The sameness hypothesis predicts these effects, whereas the QR hypothesis does not.
- (2) Experiments 3 and 4 demonstrate that in the case of definite NPs, there are better ways available to express the intended meaning: either by the insertion of *also* (Experiment 3) or by use of *the same* (rather than just *the*; Experiment 4). The availability of these more appropriate ways to express the meaning leads to a decrease in the acceptability of the *the* conditions with ellipsis and with the same verbs. Crucially, the same pressure is not there in the *every* conditions: for the case of *the same*, there is no form **every same* and hence no competition from this alternative way of expressing the meaning; and for the situation with *also*, this option is available with *every*, but it turns out that addition of *also* actually degrades (rather than improving) the *every* conditions.
- (3) Experiment 5 shows that HKV’s explanation for the difference between the *the* and *every* conditions is unlikely to be correct, because the difference disappears when the sentences are presented in supportive contexts. These context effects are not predicted by the QR hypothesis, but they are predicted by the

sameness hypothesis: indeed, there is independent evidence that the pressure to highlight the ‘sameness’ disappears when the context establishes some other connection between the events.

2 EXPERIMENT 1

In the ACD conditions in HKV’s experiments, the intended meaning of the elided material is that of the matrix verb, e.g. *was* in (5a)/(5b) is understood as the verb complex *was negotiating with*. But in each of HKV’s experiments, the control (full relative clause) conditions contained a verb that is different from the matrix verb. For example, in (6), the relative clause control verb complex is *funded*. HKV attribute the interaction that they observe to syntactic differences among the conditions, but they don’t rule out the possibility that the critical effect could be due to other differences among the conditions.

HKV justify not using the same verb (as in the matrix clause) in the no-ellipsis conditions as follows, ‘To prevent possible interference due to anaphoric down-stressing, which is subject to licensing constraints that are very similar to ellipsis [see Tancredi 1992 among others], we chose lexical verbs that were different from the matrix verb’ (HKV, p. 166). HKV do not elaborate further on the nature of the possible interference, and so we can only attempt to reconstruct their reasoning. HKV apparently assume that the conditions on downstressing/deaccenting the embedded verb *read* in an example like *Bill read the/every book that John read* are such that deaccenting requires the presence of a full VP whose logical form is formally identical to that of the deaccented one. (This is a controversial assumption in any case—see Rooth 1992.) It therefore follows that deaccenting requires *read t* not only in the relative clause but also in the matrix clause. From there, we reconstruct HKV’s reasoning as follows:

In normal speech, a sentence that has the same verb in the relative clause as in the matrix clause is usually produced with deaccented prosody on the relative clause verb. Thus, in a sentence like *Bill read the/every book that John read*, the second occurrence of *read* is normally deaccented. Given that such a verb is generally produced with deaccented prosody, when reading the sentence comprehenders will most likely adopt this prosody (e.g. Fodor 1998). But then the processor would have to find a way to license that prosody. From here on, the story is the same as for ACD: for the case of *every book that...* (but not for the case of *the book that...*) QR has

already applied, resulting in a matrix VP of the form *read t*. As a result, in the *the* conditions the processor will need to go back and perform QR in order to license the relevant prosody. Hence, a control condition with the same verb in its relative clause is not a good control, because the QR hypothesis predicts the same effect for this alternative.

This reasoning is problematic in at least two ways. First, a *speaker* knows to deaccent because she/he knows the intended meaning (and/or LF). But a *comprehender* doesn't know this. Indeed, this is exactly what the comprehender is trying to find out. Second, these materials are read, and so 'deaccenting' does not have the same status as ellipsis. Ellipsis is part of the input stimulus, such that the processor must find an antecedent for the ellided phrase. Deaccenting, on the other hand, is not given in read stimuli. Why, then, would the processor initially posit deaccented prosody for the case with the same verb on its first pass through? Under HKV's own assumptions, it should not. In particular, because HKV critically assume that QR applies only when necessary, QR can't have applied when initially processing a relative clause with the same verb as the matrix verb. Thus the conditions for deaccenting aren't met, and consequently the processor would assign the same prosody as it would for a different verb. Because the sentence is interpretable with this prosody, the processor would have no reason to go back and reanalyze with QR and deaccenting.¹⁰

Hence, under what we assume to be HKV's own assumptions, a control relative clause with the same verb will not require QR in the

¹⁰ A possible response to this reasoning is that maybe the processor can nonetheless detect that there is enough of a similarity between the two meanings that it would supply the deaccented prosody, and then it would have to go back to license this prosody. But if 'sufficient similarity' (without identity of LF) is enough to allow the processor to use deaccented prosody, then the conditions for deaccenting do not require identical LFs and again there would be no need for the processor to reanalyze.

Indeed there is good evidence that the conditions for deaccenting are much weaker than identity for ellipsis, and that deaccenting can be licensed by contextually understood material which is not overt (e.g. Rooth 1992). But HKV's rationale for not controlling for the same verb assumes that LF identity is required.

Finally, one might counter our reasoning by suggesting that use of the same verb is enough to trigger deaccenting: the processor immediately posits the deaccented prosody simply in virtue of the verb being repeated. But such a claim would need to be demonstrated. After all, there are many cases of use of the same verb where indeed deaccenting does not take place, as in (i)

- (i) John read the book that the woman who had read *Crime and Punishment* recommended.

Here the second occurrence of *read* is not deaccented. If, then, the processor in reading materials like this always provided deaccented prosody as a default just because a verb is repeated, it would have to retreat from that in a case like (i). There is, then, no reason to believe that the processor supplies a deaccented prosody in the control case for ACD, and so no reason to dismiss this as a crucial control.

the conditions. Consequently, any advantage of *every* over *the* in a control relative clause with the same verb as in the matrix clause cannot be attributed to the processor having to reanalyze the *the* condition using an extra QR step.¹¹

Given that there is no reason that the same verb condition would force QR, it is crucial to use this condition to control for (at least) two potential confounds. The first is plausibility, which is especially suggestive in some examples, such as (11) (HKV's 11, from their Experiment 1 items), which have very different meanings:

(11)

- a. ACD ellipsis conditions: The anxious old farmer was rounding up the / every cow that the ignorant farm hand was.
- b. Control verb conditions: The anxious old farmer was rounding up the / every cow that the ignorant farm hand released.

Examples like this suffer from two problems. First *round up* requires a plural or collective object; *rounding up the cow* makes little sense. Second, it is strange for one person to round up a group that someone else already had. The first of these problems is shared by both versions of *the* (with ellipsis and with a different verb), and the second is shared by both *the* and *every* with ellipsis. But the condition with *the* and with ellipsis (or, with use of the same verb) has both problems, making this less acceptable than the others.

The second potential confound is suggested by the sameness hypothesis in (7), according to which the advantage of *every* over *the* is not restricted to ellipsis (although it may be stronger in ellipsis), but has to do with the pressure to highlight the 'sameness' of the verb meanings by using an alternative expression in the *the* conditions (e.g. adding *also* or using *the same* instead of just *the*). Thus the interaction that HKV observed could be due to either of these factors, and not differences in QR.

In order to evaluate the potential contributions of these two confounds to the effects observed by HKV, we added an additional control condition to their materials, using the full-verb relative clause form of the ellipsis materials. The sameness hypothesis then predicts that the pressure to use a competing form in the *the* condition should be present in both ellipsis and non-ellipsis cases. Thus the sameness hypothesis predicts an interaction between determiner type (*the*, *every*) and relative

¹¹ Moreover, foreshadowing the results of Experiment 1, the same-verb relative clause controls show a similar interaction as in HKV's original Experiment 1, so there is clearly an effect to explain. In other words, the effect of ellipsis (or the interaction between determiner and the presence of ellipsis) does not disappear, as HKV's assumptions about deaccenting would suggest.

clause verb (same verb, different verb). In contrast, the QR hypothesis predicts no such interaction. In addition, the sameness hypothesis is also consistent with an interaction between determiner type (*the, every*) and relative clause verb (same-verb, ellipsis) if, for example, the pressure for use of *also* or *same* is greater in ellipsis.

2.1 Procedure

As discussed in the introduction, the only reliable interactions between quantifier and ellipsis in HKV's data were in the off-line acceptability rating data. We therefore used an acceptability rating paradigm in the current studies.

Similar to HKV's materials, the only differences among the conditions were the determiner (*the / every*) and the verb in the relative clause. Thus any rating differences among the conditions must be due to these differences. Participants received the following instructions:

Instructions:

This survey consists of 60 sentences, each followed by a question, and a naturalness rating of part of the sentence in an empty context. Read each target sentence, rate how natural the target sentence sounds, and answer the question immediately following.

The naturalness/acceptability ratings were presented as five choices corresponding to five radio buttons, with the responses later converted to numbers from 1 to 5 as follows:

- 1: Extremely unnatural;
- 2: Somewhat unnatural;
- 3: Possible;
- 4: Somewhat natural; and
- 5: Extremely natural.

The experiment took approximately 20 minutes to complete.

2.2 Participants

We posted surveys for 90 workers on Amazon.com's Mechanical Turk using the Turkolizer software from Gibson *et al.* (2011). All participants were paid for their participation. Participants were asked to indicate their native language, but payment was not contingent on their responses to this question.

2.3 Design and materials

The materials consisted of 60 sets of sentences edited directly from HKV's materials from Experiment 1. HKV had included additional sentence-final material in each item (e.g. *in order to arrange for free vaccination clinics* in (5) and (6)) so that they could measure reading time difficulty following the complex NP object. In order to focus attention on the relevant part of the sentences, we omitted this sentence-final material in our items. In a 2×3 design, the determiner introducing the object NP (*the, every*) was crossed with the verb-type in the relative clause structure (ellipsis, different-verb, same-verb). The different-verb condition was taken from HKV's materials, where the relative clause verb is different from the matrix clause verb (e.g. *funded* in (6)). The same-verb condition was the edited version of HKV's ellipsis condition, where the relative clause verb is the same as the matrix clause verb (e.g. *was negotiating with* in (5)). An example item is presented in (13).

(13) The understaffed general hospital was negotiating with . . .

- | | |
|-----------------------|---|
| a. ellipsis, every: | every doctor that the non-profit medical organization was. |
| b. ellipsis, the: | the doctor that the non-profit medical organization was. |
| c. diff. verb, every: | every doctor that the non-profit medical organization funded. |
| d. diff. verb, the: | the doctor that the non-profit medical organization funded. |
| e. same verb, every: | every doctor that the non-profit medical organization was negotiating with. |
| f. same verb, the: | the doctor that the non-profit medical organization was negotiating with. |

Each item was associated with a comprehension question, asking about some aspect of the sentence. These were included to ensure that participants read and understood the target sentences. An example question for (13) is presented in (14).

(14) Was the general hospital understaffed? (Yes)

Correct 'yes' and 'no' responses were balanced across items such that each list had equal numbers of 'yes' and 'no' answers in each condition. There were no distractor items in the survey.

We used a Latin Square design to create 6 lists of the 60 items, each with a counterbalanced set of items from each condition.

Each participant received a different randomized order of one of the six lists. These properties apply to all further experiments presented here.

2.4 Results

Only data from native English speakers from the United States were analyzed. We also excluded participants with less than 75% accuracy on the questions. These two exclusion criteria left data from 86 participants in Experiment 1 that we used in the analyses below.

We fit a mixed-effects linear model predicting z -transformed acceptability ratings (means and standard deviations estimated within participants) from relative clause verb (3-levels, simple coding, centered), and determiner (2-levels, sum-coded, centered). The relative clause verb was simply coded because we wanted to test both whether the same-verb condition differed from the different-verb condition, and whether the same-verb condition differed from the ellipsis condition.

Analyses reported here were conducted with the `lme4` package (Bates *et al.* 2008) for the statistical language R (R Core Development Team 2008). Recent results have shown that including only random intercepts in linear mixed-effects regressions can be anti-conservative, so we also include random slopes for all fixed effects grouped by participants and items in our model (Barr *et al.* 2013). Significance (p) values were estimated from the t -statistic. Because of the large number of data points in all of our experiments, reliable t -values are those that are larger than approximately 1.96.

The results of the model are summarized in Table 1 of section A of the supplementary data. See section B of the supplementary data for a summary of the mean ratings for all experiments presented here. Critically, the difference between the *every* condition and the *the* condition was smaller in the same verb condition than in the ellipsis condition (mean rating 3.73 (*every*) – 3.51 (*the*) = 0.22 for same verb condition; average rating 3.17 – 2.83 = 0.34 for the ellipsis condition). And the difference between the *every* condition and the *the* condition was larger for the same verb condition than for the different verb condition (mean rating 3.73 (*every*) – 3.51 (*the*) = 0.22 for same verb condition; average rating 4.21 – 4.20 = 0.01 for the different verb condition). These interactions were reliable ($t = -1.98$ for the interaction between determiner and verb type for the contrast between the same verb v. ellipsis; $t = 3.05$ for the contrast between determiner and verb type for the contrast between same verb v. different verb). Figure 1 depicts z -scored condition means and 95% confidence intervals based on the standard error of the condition mean.

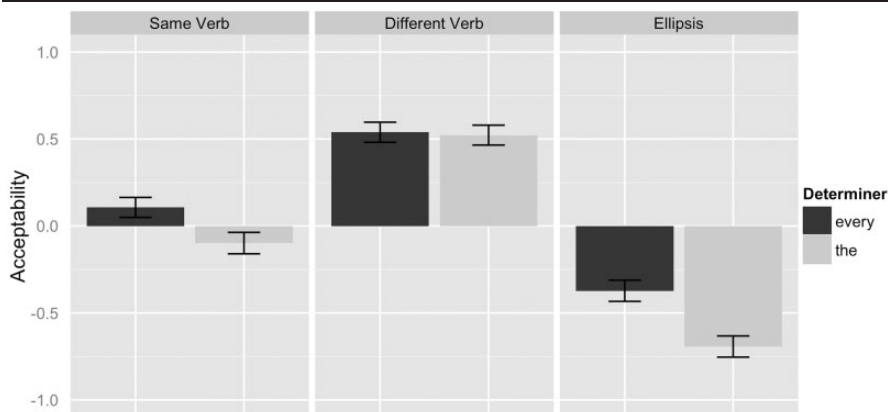


Figure 1 Z-transformed acceptability ratings for Experiment 1, crossing determiner (*the*, *every*) and relative clause verb (ellipsis, different-verb, same-verb). Error bars show 95% confidence intervals.

2.5 Discussion

First, replicating HKV's results, we found an interaction between determiner and verb type (ellipsis, different-verb). As discussed above, the sameness hypothesis, but not the QR hypothesis, predicts this interaction in whole-sentence acceptability ratings. Nevertheless, we observe such interactions in this experiment and in every further experiment to be reported here.

Second, we found similar results in the analysis of the same-verb version of HKV's materials, in the form of an interaction between determiner and verb type where verb type is restricted to the ellipsis and same-verb. This result suggests that HKV's critical interaction is still present when the plausibility and the 'sameness' confound are controlled, for, but the effect is reduced.

Finally, as predicted by the sameness hypothesis, we found an interaction between determiner and verb type where verb type is restricted to same-verb and different-verb. This interaction suggests that some of HKV's critical effect has nothing to do with ellipsis and hence nothing to do with QR. Moreover, as was speculated above, while there might be pressure to use a competing form in the *the* (but not in the *every*) condition when the same verb is used across the two clauses, this pressure could well be stronger when the second verb is elided as ellipsis directly draws attention to the similarity of the meaning in the two clauses. Thus the results of Experiment 1 are consistent with our sameness hypothesis.

3 EXPERIMENT 2

In Experiment 2, we evaluated HKV's Experiment 2 materials (small ellipsis conditions), similar to Experiment 1.

3.1 Procedure

We used the same acceptability-rating procedure as in Experiment 1.

3.2 Participants

We posted surveys for 60 workers on Amazon.com's Mechanical Turk. None of the participants took part in Experiment 1.

3.3 Design and materials

The materials consisted of 60 sets of sentences edited directly from HKV's 'small ellipsis' and full relative-clause verb conditions in their Experiment 2. An example is provided in (15). (15a)–(15d) are edited directly from HKV's materials, with the padding at the end of the sentences removed, as in Experiment 1; (15e) and (15f) are the same-verb plausibility controls.

(15) The doctor was reluctant to treat . . .

- | | |
|-----------------------|---|
| a. ellipsis, every: | every patient that the recently hired nurse did. |
| b. ellipsis, the: | the patient that the recently hired nurse did. |
| c. diff. verb, every: | every patient that the recently hired nurse admitted. |
| d. diff. verb, the: | the patient that the recently hired nurse admitted. |
| e. same verb, every: | every patient that the recently hired nurse treated. |
| f. same verb, the: | the patient that the recently hired nurse treated. |

Each item was associated with a comprehension question, asking about some aspect of the sentence. An example question for (15) is presented in (16).

(16) Was the doctor reluctant to treat a patient? (Yes)

As in Experiment 1, the sameness hypothesis predicts that the *the* condition should be rated as less acceptable than the *every* condition even in materials without ellipsis, because there is a competing form *the same* when the same verb is used, whereas there is no competing form **every*

same.¹² Thus, an interaction is predicted between determiner and verb type where verb type is restricted to same-verb and different-verb.

3.4 Results

Only data from native English speakers from the United States were analyzed. We also excluded participants with less than 75% accuracy on the questions. These two exclusion criteria left data from 48 participants in Experiment 2 that we used in the analyses below.¹³

As in Experiment 1, we fit a mixed-effects linear model predicting z -transformed acceptability ratings (means and standard deviations estimated within subjects) from relative clause verb (3-levels, simple coding, centered), and determiner (2-levels, sum-coded, centered). The model included random intercepts for participants and items as well as random slopes for all fixed effects grouped by participant and item.

The results are summarized in Table 2 of section A of the supplementary information. Critically, the difference between *every* and *the* was greater in the same verb condition ($3.89 - 3.55 = 0.34$) than in the different verb condition ($4.18 - 4.05 = 0.13$). This interaction was reliable ($t = -2.02$). There was no significant interaction, however, between determiner and verb type for the contrast between the same verb condition and the ellipsis condition ($t = 0.09$). Figure 2 depicts z -scored condition means and 95% confidence intervals.

3.5 Discussion

As in Experiment 1, we replicated HKV's result: an interaction between determiner and verb type (ellipsis, different-verb). However, the interaction between determiner and verb type where verb type is restricted to the ellipsis and same-verb was not reliable. Most critically for the sameness hypothesis, we found a reliable interaction between determiner and verb type where verb type is restricted to same-verb and different-verb, with no ellipsis at all. As for Experiment 1, this

¹² In Experiment 1, there was an additional potential competing form, with *also* for the *the* conditions. This alternative is not available in the materials in Experiment 2, because of the presence of a higher level predicate (e.g. *was reluctant to* in (15)). With both *the* and *every*, and with or without ellipsis, there is a preference to interpret the object NP *de dicto*, that is, with scope under *reluctant*. Given the presence of this predicate, it is not possible to insert *also* here, because the presupposition of *also* is not met: the use of *also* induces a presupposition that there is some other individual that treated the (or each) relevant individual, and in the case where the main predicate is *is reluctant to* nothing in the main clause introduces material that would satisfy this presupposition. Most of the upper predicates in HKV's Experiment 2 materials have this property. Hence, the competing form with *also* is unavailable in both the *the* and the *every* condition.

¹³ Of the 12 surveys that we did not analyze in this experiment, most were from a single individual who did not read the instructions, and filled out multiple versions of the survey.

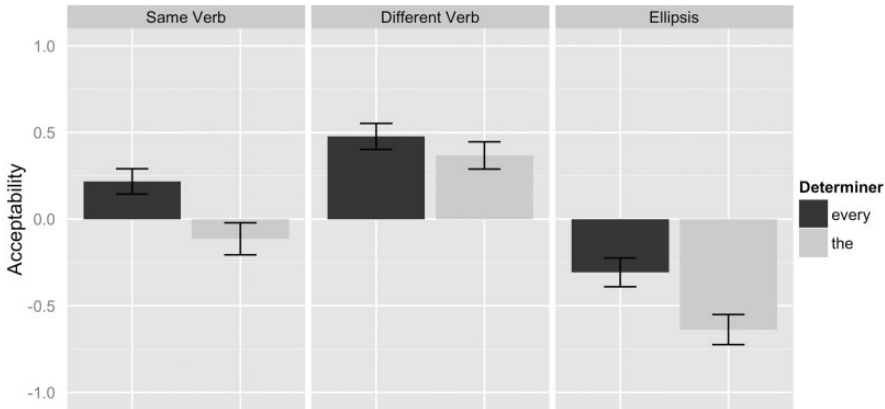


Figure 2 Z-transformed acceptability ratings for Experiment 2, crossing determiner (*the*, *every*) and relative clause verb (ellipsis, different-verb, same-verb). Error bars show 95% confidence intervals.

interaction suggests that HKV's critical effect may be driven by differences among the conditions having nothing to do with ellipsis and hence nothing to do with QR.

These results suggest that HKV's critical interaction in Experiment 2 (small ellipsis) is not present—or is at least greatly reduced—when plausibility (and other potential confounds) is controlled for by using the same verb. Exactly why there was still an effect for ellipsis v. same-verb in Experiment 1 but not in Experiment 2 is unclear. It may simply be that there is a small effect to be found, but we didn't find it in this experiment. This issue is not critical to disentangling the sameness v. QR hypotheses, so we leave this issue open.

4 DIRECT EVALUATIONS OF THE SAMENESS HYPOTHESIS

In the remaining three experiments, we evaluate the sameness hypothesis in more detail. In particular, in the null context there is a strong pragmatic pressure to call attention to the 'sameness' of the verbs across the two clauses in examples like the ones HKV investigated, which can be achieved either by using the determiner *the same* or by inserting *also*. This pressure exists in the *the* condition but not in the *every* condition because the potentially competing form either does not exist or is dispreferred with *every*. In particular, with respect to *the same*, there is no pressure to use this form for the *every* conditions because there is no

form **every same*. Why *also* would be preferred in the null context for the *the* but not for the *every* conditions remains an open question, given that both *the* and *every* can occur with *also*, although below we do offer some speculation as to why *every* does not show the pressure to use *also*. In any case, our results below show that—for the materials from Experiment 1 above—insertion of *also* is indeed preferred with *the* and is dispreferred with *every*. We leave it open here as to why this is so, but the fact that *also* in the ACD case lowers the acceptability with *every* and raises it with *the* is consistent with our hypothesis.

Why should there be pressure to use *also* or *the same* with *the* (in cases where the meaning and presuppositions make these forms available)? One possibility is that the use of these forms helps satisfy the presupposition(s) associated with *the*. Definite NPs are normally used to refer to contextually salient or unique entities (or, in some cases, entities easily inferable from common knowledge such as *the sun*). Because HKV's materials were presented with no context, a comprehender would have to come up with a scenario to establish the referent of the definite NP. The use of *also* or *the same* would help in doing so. For example, in *The doctor treated the same patient that the nurse did*, the presence of *same* immediately allows the listener to infer that someone else treated the relevant patient since *same* requires that the verb (or understood verb) must be *treat* (without *same*, any later verb would be possible), and this in turn provides information about the relevant referent and helps set up the necessary background context. A second (not mutually exclusive) possibility is that—unless some independent connection between the events can be established—there is pressure to draw attention to the sameness of the events. Below we offer some brief speculation as to why the *every* condition allows for an independent connection to be established and hence does not exhibit the same pressure.

To directly test the sameness hypothesis, in Experiment 3 we first evaluate how the presence of *also* affects the acceptability of the HKV materials. We find that the presence of *also* lowers the acceptability of the *every* examples with ACD, but increases the acceptability of the *the* examples with ACD. In fact, the advantage of *every* over *the* entirely disappears with ACD when *also* is present in the examples.¹⁴

¹⁴ The word *also* introduces a presupposition, and some accounts of how *also*'s presupposition is satisfied would predict that QR must happen at the site where *also* is first encountered. This, however, does not alter the prediction of the QR hypothesis account for ACD conditions which include *also*: the QR hypothesis still predicts an advantage for *every* over *the*. In particular, the presence of *also* just moves the location where QR occurs in the *the* case earlier in the input: with *also* present, QR would occur when *also* is encountered rather than at the ACD site. But as in the case without *also*, QR will already have applied in the *every* condition at this point. Thus the

This pattern is consistent with our hypothesis whereby the effect is driven by a pressure in the null context to use a form which highlights the ‘sameness’. Second, in Experiment 4, we show that the inclusion of *same* in the *the* / ACD examples greatly improves their acceptability. In contrast, it is not possible to use *same* with *every* (*the doctor - the same doctor; every doctor - *every same doctor*). Although the fact that *the same* is rated as highly as *every* is consistent with HKV’s hypothesis—under the assumption that *the same* also undergoes QR—it is as predicted by the sameness hypothesis. Finally, we show in Experiment 5 that the difference between the *the* and *every* conditions disappears in a supportive context. This pattern is again predicted by the sameness hypothesis, because (a) the presuppositions of *the* are satisfied in the context, and (b) there is an independently established connection between the two events and so there is no pressure to use *also* or *the same* in the *the* condition. In contrast, the QR hypothesis cannot explain these contextual effects. Together, these experiments support the sameness hypothesis over the QR hypothesis.

5 EXPERIMENT 3

In Experiment 3, we evaluate how the presence of the word *also* in items like those used in HKV’s Experiment 1 affects the acceptability of the ACD examples.

5.1 Procedure

We used the same acceptability-rating procedure as in Experiments 1 and 2.

5.2 Participants

We posted surveys for 60 workers on Amazon.com’s Mechanical Turk. None of the participants took part in Experiments 1 or 2.

5.3 Design and materials

The materials consisted of 60 sets of sentences in which 4 of the conditions were taken directly from the Experiment 1 materials: the ellipsis and same-verb conditions (which were themselves edited directly from

QR hypothesis makes the same prediction for ACD materials with *also*: an advantage of the *every* conditions over the *the* condition. For more details on the interaction of antecedent containment and *also* see Jacobson (2009).

HKV's Experiment 1 materials). In addition, we included a condition with ACD ellipsis and the word *also* as in (17e)–(17f):

(17) The understaffed general hospital was negotiating with . . .

- | | |
|--------------------------|---|
| a. ellipsis, every: | every doctor that the non-profit medical organization was. |
| b. ellipsis, the: | the doctor that the non-profit medical organization was. |
| c. same verb, every: | every doctor that the non-profit medical organization was negotiating with. |
| d. same verb, the: | the doctor that the non-profit medical organization was negotiating with. |
| e. ellipsis+also, every: | every doctor that the non-profit medical organization also was. |
| f. ellipsis+also, the: | the doctor that the non-profit medical organization also was. |

As in previous experiments, each item was associated with a comprehension question, asking about some aspect of the sentence.

5.4 Results

Only data from native English speakers from the United States were analyzed. We also excluded participants with less than 75% accuracy on the questions. These two exclusion criteria left data from 56 participants in Experiment 3 that we used in the analyses below.

As in previous experiments, we fit a mixed-effects linear model predicting z -transformed acceptability ratings (means and standard deviations estimated within subjects) from relative clause verb (3-levels, simple coding, centered), and determiner (2-levels, sum-coded, centered). The model included random intercepts for participants and items as well as random slopes for all fixed effects grouped by item and all main fixed effects grouped by participant.¹⁵

The results are summarized in Table 3 of section A of the supplementary data. The difference in mean rating between *every* and *the* was greater in the ellipsis condition ($3.05 - 2.59 = 0.46$) than in the full verb condition ($4.30 - 4.22 = 0.08$), which was a reliable interaction ($t = -5.59$). Critically, the difference in mean rating between *every* and *the* was also greater in the ellipsis condition ($3.05 - 2.59 = 0.46$) than in the *also* condition ($2.86 - 2.80 = 0.06$; significant at $t = -5.12$).

¹⁵ Models with random slopes for interactions grouped by participant did not converge.

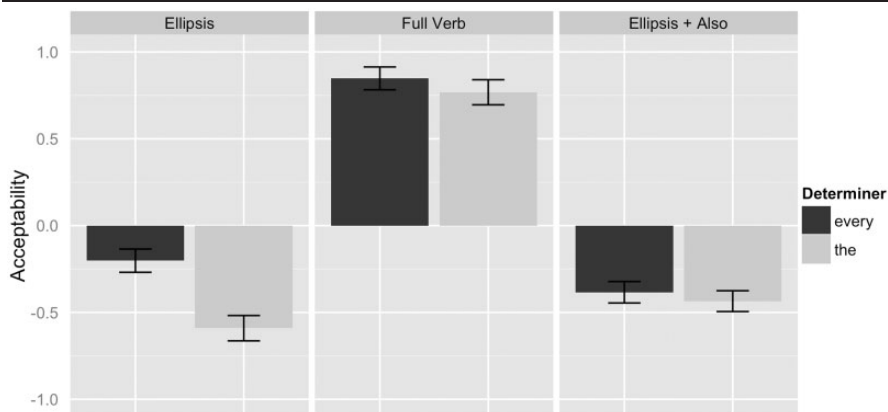


Figure 3 Z-transformed acceptability ratings for Experiment 3, crossing determiner (*the*, *every*) and relative clause verb (ellipsis, full-verb, ellipsis+also). Error bars show 95% confidence intervals.

Figure 3 depicts z -transformed condition means and 95% confidence intervals.

5.5 Discussion

As in Experiment 1, we found an interaction between determiner and verb type, using plausibility-controlled versions of HKV's materials. Second, we found an interaction between determiner and the presence of *also*, such that the presence of *also* lowered the acceptability of the *every* / ACD examples, but increased the acceptability of the *the* / ACD examples. This result supports the sameness hypothesis. The fact that *also* lowers the acceptability of the *every* cases explains why there is no pressure to use *also* in these conditions. In other words, there exist better ways to express the target meaning in the / ACD condition, but this is not true for the case of *every*.

As pointed out by an anonymous referee, it appears to be a puzzle for the sameness hypothesis that the presence of *also* improves the cases with *the* but not those with *every*. We speculate that the reason that *also* is not needed with *every* is that it is easy for speakers to construct what we might call a 'copycat' interpretation for these cases, whereby there is a causal connection between the events.¹⁶ This claim can be tested as follows. Consider the following two sentence frames, with respect to a potential causal connection between the clauses in each, as in (20):

¹⁶ We are grateful to Geoff Pullum for pointing out to us the 'copycat' interpretation.

- (18) John read every book that Mary did. Mary read Crime and Punishment.
- (19) John read the book that Mary did. Mary read Crime and Punishment.
- (20) John read Crime and Punishment because Mary read Crime and Punishment.

We predict that people will judge (20) to be more likely to be true in the (18) scenario (which establishes the causal connection by use of the word *every*) than in the scenario in (19). Initial support for this hypothesis is provided in Jacobson & Gibson (2014). Thus the sameness hypothesis, supplemented with the greater availability of a causal connection for *every* than for *the* leads us to expect a greater pressure to insert *also* in the *the* condition than in the *every* condition.

6 EXPERIMENT 4

In Experiment 4, we evaluated how the presence of the word *same* in items like those used in HKV's Experiment 1 affects the acceptability of the ACD examples. Whereas the word *same* can be added following the determiner *the* (e.g. *the same doctor*), it is not possible to include it following *every* (e.g. **every same doctor*). Thus, the presence of *same* increasing the acceptability of the *the* / ACD examples would be consistent with the sameness hypothesis. This pattern would also be consistent with the QR hypothesis under the assumption that *the same* also undergoes QR. This experiment therefore does not attempt to distinguish the two alternatives, but it is still important to evaluate this prediction of both accounts.

6.1 Procedure

We used the same acceptability-rating procedure as in previous experiments.

6.2 Participants

We posted surveys for 60 workers on Amazon.com's Mechanical Turk. None of the participants took part in previous experiments.

6.3 *Design and materials*

The materials consisted of 18 sets of sentences from 20 sets of sentences that we constructed for Experiment 5, where we created sentences which were plausible in supportive contexts. These items were based loosely on the items from HKV's Experiment 2. The current experiment crossed determiner (*the, every, the same*) with verb type (ellipsis, full verb). An example is provided in (21):

(21) The choreographer evaluated

- a. ellipsis, every: every ballerina that the lead dancer did.
- b. ellipsis, the: the ballerina that the lead dancer did.
- c. ellipsis, the same: the same ballerina that the lead dancer did.
- d. full verb, every: every ballerina that the lead dancer evaluated.
- e. full verb, the: the ballerina that the lead dancer evaluated.
- f. full verb, the same: the same ballerina that the lead dancer evaluated.

As in previous experiments, each item was associated with a comprehension question, asking about some aspect of the sentence. The full sets of materials for Experiments 4, 5 and 6 are provided in section C of the supplementary data.

6.4 *Results*

Only data from native English speakers from the United States were analyzed. We also excluded participants with less than 75% accuracy on the questions. These two exclusion criteria left data from 57 participants in Experiment 3 that we used in the analyses below.

As in previous experiments, we fit a mixed-effects linear model predicting z -transformed acceptability ratings (means and standard deviations estimated within subjects) from relative clause verb (2-levels, sum-coded, centered), and determiner (3-levels, Helmert-coded, centered). The Helmert contrast compared (a) *the same* v. *every* (which were not predicted to differ according to the pragmatic explanation); and (b) *the same* and *every* together v. *the*. The model included random intercepts for participants and items as well as random slopes for all fixed effects grouped by participant and item.

The results are summarized in Table 4 of section A of the supplementary data. Critically, the difference in mean rating between the full verb condition and the ellipsis condition was larger for the *the* condition (mean rating $3.53 - 2.89 = 0.64$) than for the difference across the other two conditions ($3.91 - 3.72 = 0.19$). This interaction between

determiner and verb type (ellipsis, full verb), where the *the* condition was compared with the *every* and *the same* conditions together, was reliable ($t = 4.34$). But there was no reliable interaction between determiner and verb type for just the *every* and *the same* values of this factor ($t = 0.50$). Figure 4 depicts z -scored condition means and 95% confidence intervals.

6.5 Discussion

There were two critical results in this experiment. First, there was a reliable interaction between determiner and relative clause verb (ellipsis, full-verb), where the *the* condition was compared with the *every* and *the same* conditions together. This replicated previous results in this article. Second, as predicted by the sameness hypothesis, there was no interaction between determiner and relative clause verb when comparing just the *every* and *the same* conditions. Under the sameness hypothesis, there is general pressure to highlight the ‘sameness’ of the actions in the two clauses, but because the form **every same* is not possible, there is no such pressure in the *every* condition. As noted earlier, the fact that *the same* is rated as highly as *every* is also consistent with the QR hypothesis. On their own then, the results of this experiment do not distinguish between the QR hypothesis and the sameness hypothesis, but taken together with the results of previous experiments, we have evidence for the sameness hypothesis over the QR hypothesis.

7 EXPERIMENT 5

According to the sameness hypothesis, part of the difficulty with the *the* / ACD condition has to do with it being presented in a null context, for two reasons. First, as discussed above, the pressure to highlight the ‘sameness’ of the action can be greatly ameliorated if the context supplies some other connection between the two events. Second, use of *the* requires a previously established contextually salient referent (or one that can easily be inferred from background assumptions or world knowledge). In the null context, no such referent exists, but the addition of *the same* or *also* helps the processor construct such a referent. In a supportive context, the advantage of *every* over *the* with ACD should disappear, according to the sameness hypothesis. This prediction contrasts with that of the QR hypothesis, which predicts no change in complexity in supportive contexts: regardless of context, the processor

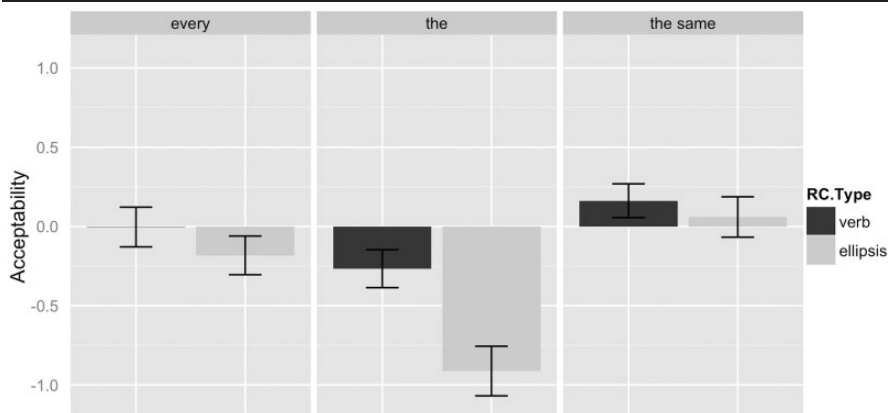


Figure 4 Z-transformed acceptability ratings for Experiment 4, crossing determiner (*the*, *every*, *the same*) and relative clause verb (ellipsis, full-verb). Error bars show 95% confidence intervals.

will always have to go back and perform QR with *the* and not with *every*. Experiment 5 tests these predictions.

7.1 Procedure

This experiment was run in two versions. Participants in the null-context version of the experiment received the same instructions as for previous experiments: to simply rate the acceptability of a sentence. Participants in the supportive-context version received the following instructions:

Instructions

1. Read the context.
2. Read the target sentence.
3. Rate how natural the target sentence sounds in the context.
4. Answer the question immediately following.

As in previous experiments, the naturalness/acceptability ratings were presented as five choices corresponding to five radio buttons, with the responses later converted to numbers from 1 to 5. The experiment took approximately 20 minutes to complete for the null-context version, and 30 minutes for the supportive-context version.

7.2 Participants

We posted surveys for 120 workers on Amazon.com's Mechanical Turk, 60 for each version.

7.3 Design and materials

The materials consisted of 20 sets of sentences based loosely on the items from HKV's Experiment 2. Each test item had eight conditions, in a $2 \times 2 \times 2$ design, crossing the determiner introducing the object NP (*the, every*), verb type (ellipsis, full-verb) and context (null, supportive). An example of a supportive context for the *the* conditions is presented in (22a). An example of a supportive context for the *every* conditions is presented in (22b). These contexts are minimally different, such that two relevant entities are introduced for the *the* conditions, whereas several relevant entities are introduced for the *every* conditions. The form of the target items is given in (22).¹⁷ See Appendix D for a full list of the materials.

(22)

- a. Supportive context for *the* conditions (components that differ with the *every* conditions are underlined):

Mary: At the dance school two ballerinas auditioned to be in the next performance. One of the ballerinas was evaluated by the director, while the other was evaluated by the lead dancer.
John: I heard that one of the ballerinas was also evaluated by the choreographer. Do you know which ballerina?

- b. Supportive context for *every* conditions (components that differ with the *the* conditions are underlined):

¹⁷ In the version of the contexts in (20), the verb that is used in the target sentence (22) is used in the passive, e.g. *One of the ballerinas was evaluated by the director, . . .*. In an earlier version of this experiment that we ran, we constructed contexts in which the verb in the target sentence was in the active voice, e.g. *The director evaluated one of the ballerinas, . . .*. A potential concern with using active voice in the contexts is that the antecedent for 'did' in the target sentences could be coming from the context sentence, and not necessarily from an ACD structure. It is independently known that transitive verb phrase ellipsis can come from a previous sentence, as in (i):

(i) Bagels, I like. Donuts I don't. (Evans 1988).

Consequently, we set up contexts where the transitive VP does not occur in the context, as in (22), so that the antecedent for 'did' in the target sentences must come from an ACD structure, as desired.

Incidentally, on some theories of antecedent-contained deletion, there are cases where the meaning of the ellipsis site is actually supplied by context and not by overt linguistic material (see e.g. Jacobson 2008). Under that view, one might argue that the context sentences themselves are enough to set up a way to supply the 'missing' material in the ACD cases. But if this is true, it undermines the entire logic of HKV's experiments: in all of the relevant cases the elided material *V-trace* could be picked up from the matrix clause and so QR would never be required. Hence given HKV's general set of assumptions, the fact that the advantage for *every* over *the* disappears with context is not predicted by their account.

Mary: At the dance school several ballerinas auditioned to be in the next performance. Some of the ballerinas were evaluated by the director, while the rest were evaluated by the lead dancer.

John: I heard that some of the ballerinas were also evaluated by the choreographer. Do you know which ballerinas?

(23) Mary: The choreographer evaluated . . .

- a. ellipsis, every: every ballerina that the lead dancer did.
- b. ellipsis, the: the ballerina that the lead dancer did.
- c. full verb, every: every ballerina that the lead dancer evaluated.
- d. full verb, the: the ballerina that the lead dancer evaluated.

Each context was associated with a comprehension question, as was each target sentence, as shown in (24) and (25), respectively. Thus, for null context materials, there was only one question, whereas for supportive context materials there were two comprehension questions.

(24) Did some ballerinas audition for a performance? (Yes)

(25) Did the choreographer evaluate the ballerinas that the lead dancer evaluated? (Yes)

In addition to the target materials, there were 32 distractor items in the survey, of similar length and complexity as the targets. For the supportive-context version of the experiment, the distractor items had contexts, and two comprehension questions each. For the null-context version, the distractor items had no contexts, and there was only one comprehension question. The full sets of materials for Experiments 4, 5 and 6 are provided in section C of the supplementary data.

7.4 Results

Only data from native English speakers from the United States were analyzed. We also excluded participants with less than 75% accuracy on the questions. These two exclusion criteria left data from 46 participants for the null-context version of the experiment and 55 participants for the supportive-context version, analyzed below.¹⁸

As in previous experiments, we fit a mixed-effects linear model predicting z -transformed acceptability ratings (means and standard deviations estimated within subjects) from relative clause verb (2-levels,

¹⁸ Of the 14 surveys that we did not analyze for the null-context version, most were from a single individual who did not read the instructions, and filled out multiple versions of the survey.

sum-coded, centered), determiner (2-levels, sum-coded, centered), and context (2-levels, sum-coded, centered). The model included random intercepts for participants and items as well as random slopes for all fixed effects grouped by item and random slopes for all main between-subject fixed effects grouped by participant.¹⁹

The results are summarized in Table 5 of section A of the supplementary data. Figure 5 depicts z -scored condition means and 95% confidence intervals.

7.6 Discussion

We observed reliable main effects of determiner (*every* rated higher overall; $t = 3.57$) and relative clause verb (full verb rated higher overall; $t = 4.24$) as well as a reliable interaction between the two ($t = 3.33$) overall, across the two kinds of context. We thus replicated HKV's interaction in a new set of materials. In addition, there was also a 3-way interaction ($t = 2.36$) among determiner, relative clause verb and context, demonstrating that the interaction of determiner and relative clause verb is much stronger in the null context. Visual inspection together with additional analyses demonstrated that there was no difference between the *every* and *the* conditions with ACD in supportive contexts (mean rating 3.89 v. 3.90), as predicted by the sameness hypothesis, but in contrast to the QR hypothesis, which predicts a main effect of *every* v. *the* for ACD conditions, independent of context. Overall, the results of this experiment clearly support the sameness hypothesis over the QR hypothesis.

We conducted a further experiment, Experiment 6, to further evaluate the supportive-context predictions of the sameness hypothesis. In that experiment, we set up contexts so that definite plurals (e.g. *the ballerinas*) could be contrasted with NPs quantified by *every* (e.g. *every ballerina*). Using this design, the exact same contexts can be used to compare all target sentences. The results of this experiment were very similar to the results of Experiment 5: (1) the same interaction between determiner and ACD presence in the null context that has been seen here in previous experiments; and (2) the lack of such an interaction in supportive contexts. Critically, there was again no difference between the *every* and *the* conditions with ACD in supportive contexts, as predicted by the sameness hypothesis, but in contrast to the QR hypothesis. In order to save space in the print version of the article, we present the data from this experiment in section D of the supplementary data.

¹⁹ Models with random slopes for interactions grouped by participant did not converge.

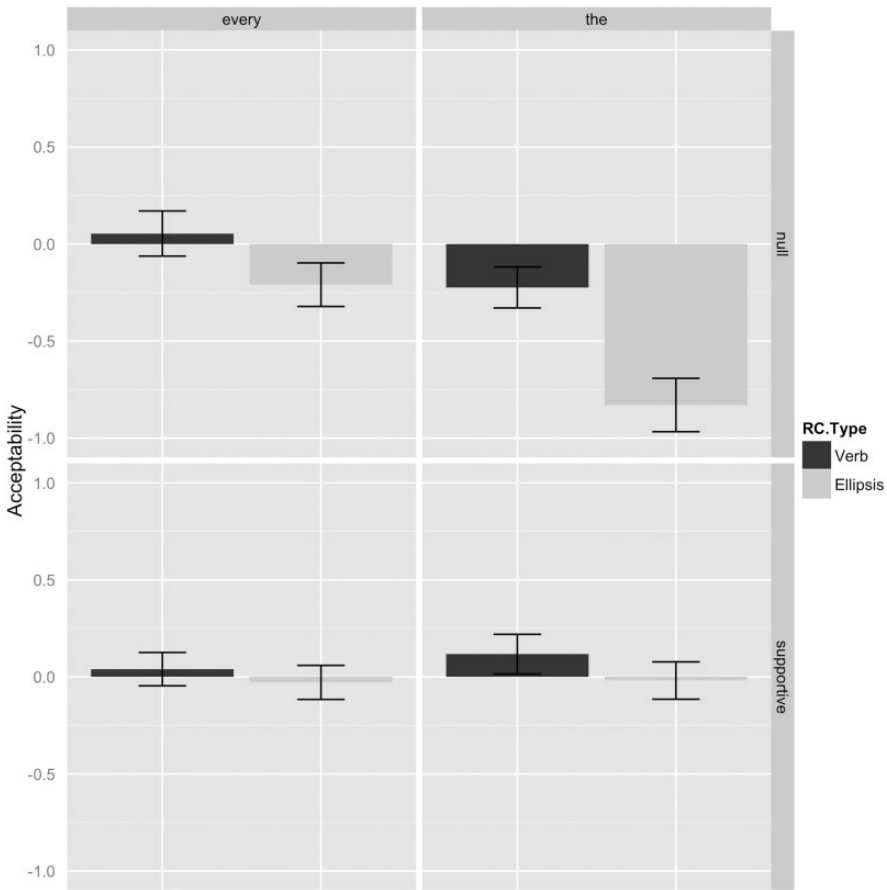


Figure 5 Z-transformed acceptability ratings for Experiment 5, crossing determiner (*the, every*), relative clause verb (ellipsis, full-verb), and context (null, supportive). Error bars show 95% confidence intervals.

8 GENERAL DISCUSSION

In other work, we have shown that HKV's critical on-line effects were not significant (Gibson *et al.* submitted). Hence, we have focused here on HKV's acceptability rating results. We presented five acceptability rating experiments that are as predicted by the sameness hypothesis, but generally not by HKV's QR hypothesis. In Experiments 1 and 2, we showed that much of the interaction between determiner and relative clause verb type which HKV observed may be explained by plausibility and/or pressure to use alternative, more appropriate, expressions than the definite NP, given that the interaction persists without ellipsis (i.e. when comparing between same-verb and different-verb conditions).

The results of Experiments 3–5 provided further evidence for the sameness hypothesis:

- (a) In Experiment 3 we showed that the presence of *also* increases the acceptability of the *the* / ACD examples, but lowers the acceptability of the *every* / ACD examples.
- (b) In Experiment 4 we showed that the inclusion of *same* in the *the* / ACD examples greatly improves their acceptability. In contrast, it is not possible to include the word *same* along with *every*. These results suggest that part of the difficulty with the *the* / ACD examples is that there is a better way to express their meaning, including the word *also* or *same*, but that this option is either not available in the *every* / ACD conditions or (in the case of *also*) lowers its acceptability and so is not a competitor.
- (c) Finally, in Experiment 5 we showed that the difference between the *the* and *every* conditions disappears in a supportive context. This result is predicted by the sameness hypothesis both because the contextually salient referent is already present (hence there is no boost from *the same* or from *also*), and because the pressure to highlight the ‘sameness’ is reduced or eliminated when a connection between the events has already been established in prior context.

Taken together, these results provide strong evidence for the sameness hypothesis as a pragmatically motivated account of the HKV rating effects. Importantly, the results are not consistent with the QR analysis proposed by HKV, because the QR analysis (i) cannot explain the advantage of *every* over *the* in the no-ellipsis conditions, when the same verb is used across the two clauses, (ii) predicts no interaction with the presence of *also* in Experiments 3, and (iii) predicts no interaction with context, as was observed in Experiment 5. Consequently, there is no empirical effect that HKV presented which is better explained by the QR hypothesis than the alternative pragmatically motivated sameness hypothesis. Hence HKV’s results do not provide any evidence for QR.

9 HKV’S EXPERIMENT 2—LARGE / SMALL ELLIPSIS

We turn now to a discussion of the large ellipsis conditions of HKV’s Experiment 2 (for further discussion see [Jacobson & Gibson 2014](#)). Their Experiment 2 had two parts. In one part, HKV examined cases where the antecedent for the ellipsis was not the main clause VP but an embedded VP: their ‘small ellipsis’ conditions, as in (26).

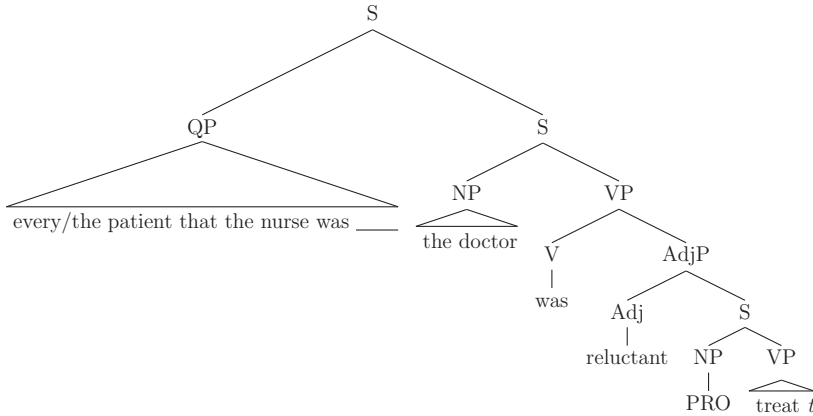


Figure 6 The structure with high QR (*de re* reading) for (26c).

- (26) a. Small ellipsis (of *treat t*): The doctor was reluctant to treat the/every patient that the recently hired nurse did after looking over the test results.
- b. Full-verb control: The doctor was reluctant to treat the/every patient that the recently hired nurse admitted after looking over the test results.

The logic of this part of the experiment is the same as in HKV Experiment 1, and they found the same contrast between *the* and *every* in off-line ratings. The second part of HKV's Experiment 2 investigated conditions where the ellipsis antecedent is the VP of the main clause—the 'large' ellipsis condition—as in (26c):

- (26) c. Large ellipsis (of *reluctant to treat t*): The doctor was reluctant to treat the/every patient that the recently hired nurse was after looking over the test results.

HKV claim that the QR hypothesis predicts that the difference between *the* and *every* should disappear in the large ellipsis conditions. This reasoning is based on two assumptions/observations. First—as known since Sag (1976)—the 'large' ellipsis (of *was reluctant to treat t* as opposed to just *treat t*) is possible only when the object is interpreted *de re*. Following Sag (and many others), HKV assume that a *de re* reading is possible only when QR raises the relevant material to the topmost clause, as shown in Figure 6 (irrelevant details—including details concerning the index on the trace—are omitted here):

Here the 'missing' material following *was* can be resolved to [_{AP} *reluctant PRO to treat t*]. This gives the *de re* reading: it can be paraphrased

as *for every/the actual patient that the nurse was reluctant to treat, the doctor was reluctant to treat that patient* (she/he may know nothing about which patients the nurse is reluctant to treat). Notice that this is a possible reading of a similar sentence in (27), without ellipsis:

- (27) The doctor was reluctant to treat every/the patient that the nurse was reluctant to treat.

Unlike the structure in Figure 6, (27) has a second *de dicto* reading, as in the scenario where the doctor believes in the nurse's judgment: if the nurse didn't want to treat a patient then the doctor reasons that she/he probably shouldn't either. Although these particular sentences require some imagination to accommodate under either reading, it is well known that in general the *de dicto* readings are easier to obtain. Sag noted, however, that the corresponding sentence with ellipsis has no *de dicto* reading, a fact which follows under the QR analysis. For under such an analysis, this reading requires the material *the/every patient that the nurse was* to be QR'ed only to the intermediate clause, as shown in Figure 7. But notice that this structure merely recreates the antecedent containment "paradox". Here there is no single constituent *reluctant to treat t* which can be used to resolve the ellipsis following *was*. Thus in the case that the ellipsis is ultimately resolved to the larger expression *reluctant to treat t* (the "large ellipsis" condition), only the *de re* wide scope reading is possible.²⁰

The next critical piece of HKV's reasoning depends on the observation noted above: that in general (when there is no ellipsis), *de dicto* readings are preferred. From this observation, they conclude that not only will the sentence processor perform QR only when needed, but it will perform only the minimal QR that it needs to do at the relevant time. Consider, then, what happens when the sentence processor encounters an object initiated by *the*. As has already been discussed, no QR will be performed here. QR will apply later only if QR is needed to resolve the ellipsis. When the sentence processor encounters *every*, it performs QR—but crucially it performs only the minimal QR—which gives the structure in Figure 7. But this is not enough to resolve the ellipsis: the sentence processor can do so only by performing another instance of QR in order to map

²⁰ One might think that the correlation between the large ellipsis reading and a *de re* reading itself argues for the QR approach, because there is a simple account of this correlation using QR (see Sag 1976). But Jacobson (1992b) showed that recasting ACD as the ellipsis of a (possibly complex) transitive verb or adjective phrase (such as *reluctant to treat*) can also account for this correlation without QR.

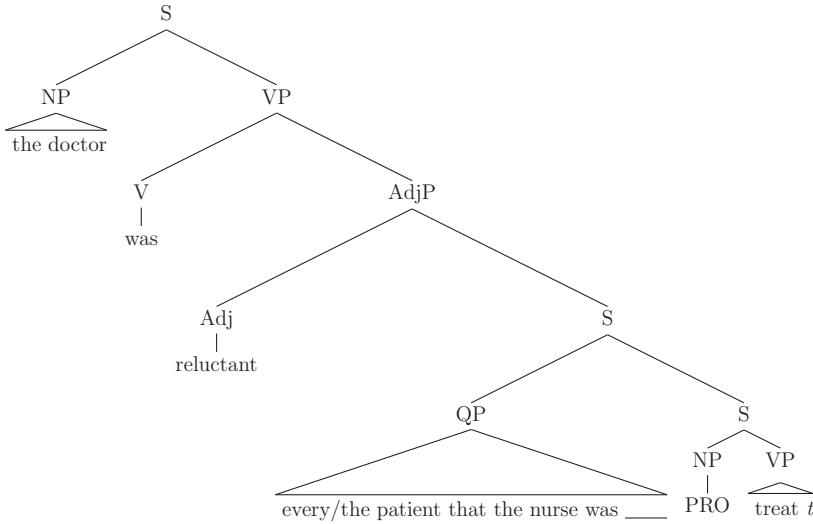


Figure 7 The structure with low QR for (26c); ellipsis fails to be resolved.

the structure in [Figure 7](#) to the one in [Figure 6](#). In other words, QR needs to apply at the ellipsis site regardless of whether the sentence processor has encountered *the* or *every* earlier on, which would therefore remove the advantage for *every*. And indeed HKV found no advantage for *every* in the large ellipsis condition.

There are flaws, however, in this explanation of the empirical data. First, there was actually a large numerical advantage for *the* over *every* in the large ellipsis conditions in HKV's data: the two were not equal. Even under the HKV interpretation of their predictions, this is unexpected. Second, contrary to their claims, the QR hypothesis actually predicts that *every* should still have an advantage over *the* in the large ellipsis conditions. Recall that HKV assume that QR will initially apply in the minimal domain; this is crucial to their story. Critically, the sentence processor cannot know in advance what meaning it is trying to compute. Thus, on encountering the ellipsis site in the *every* condition, the sentence processor cannot find an antecedent for the ellipsis and so tries an additional QR, which then gives a representation that allows the ellipsis to be resolved. Consider now what happens when the sentence processor encounters *the*. It has not performed any applications of QR. But by the minimal QR hypothesis, it will simply first perform the minimal QR, raising the *the* NP just above the *treat* clause. In other words, it will first compute the structure in [Figure 7](#), for it has no way to know in

advance that this will not be sufficient to resolve the ellipsis. (Indeed, the sentence processor can never know in advance that QR is the solution to the ellipsis resolution; one has to assume that when the sentence processor encounters an ellipsis site with no obvious antecedent it merely attempts some resolution of this, and QR is a reasonable route for it to take.) But—just as in the *every* case—the structure with the minimal QR in Figure 7 is indeed not enough to resolve the ellipsis, and so the sentence processor will perform a second application of QR. In the end, then, this is analogous to the basic (Experiment 1) case: an extra computation is needed at the point of the ellipsis for the *the* case which is not needed for the *every* case. (Here one application of QR will apply at the ellipsis site for the *every* case, and two applications will apply for the *the* case.) Thus, all other things being equal, the QR hypothesis also predicts an advantage for *every* over *the* in the large ellipsis conditions.²¹

It turns out that the sameness hypothesis as developed in the present study appears to make the same prediction as the QR hypothesis for these conditions—an advantage for *every* over *the*—because these examples are parallel to the examples in Experiment 1. In (26c), for example, the nurse and the doctor are performing the same action (for each patient, they are both reluctant to examine that patient). And the presuppositions for *the* are not set up by prior context. Consequently, there should be a pressure to use *the same* or *also* in the *the* condition (both of these are indeed possible in the *the* condition), which should lead to more complexity for the *the* condition relative to the *every* condition. However in Jacobson & Gibson (2014) we propose an explanation for why *every* loses its advantage in this condition. The essential point is that in the simple case the pressure for *also* is absent with *every* because a causal connection can be established; the types of stimuli used by HKV (and indeed necessitated by the experimental task) are such that they generally preclude establishing such a connection.

As to why the reverse result holds (*the* actually has an advantage over *every*), we suspect that the explanation for this is that all other things are not equal. A plausible account (under either theory) for the advantage of

²¹ HKV's discussion seems to assume that—on encountering the ellipsis site with *the* in the large ellipsis condition—the parser knows that it needs to perform the maximal QR. Thus they say 'When the non-local ACD site is hosted by a definite DP, however, the parser can determine at the point where QR is triggered, that is when the parser encounters the ACD site marked by *was* also how far the object DP has to be moved. Thus only one instance of reanalysis is necessary' (HKV, p. 182; fn. 45). But under HKV's own assumptions, it doesn't seem possible that the parser can—at the relevant point—determine 'how far the object DP has to be moved' because it does not know what meaning it is trying to compute. If there is some algorithm available to the processor allowing it to determine that maximal QR is needed, this algorithm needs to be spelled out.

the over *every* is that it is much easier to get the wide scope (*de re*) reading with *the* than with *every*. This preference could well be strong enough to override any advantage that *every* might have over *the* for other reasons. This of course is speculation at this point. But given that neither account by itself predicts the observed advantage for *the* over *every*, nothing at this point can be concluded from the large ellipsis conditions in HKV's Experiment 2.

10 CONCLUDING REMARKS

It is worth putting into context HKV's attempt to find evidence for empty elements and/or movement in linguistic representations. There is a long history of attempts to find evidence that would argue for or against such hypotheses. For example, Pickering & Barry (1991) provided data that seemed to support linguistic theories that lacked empty elements in long-distance dependencies. But Gibson & Hickok (1993) showed that Pickering & Barry's data could be accommodated by a theory that contains empty elements, under slightly different parsing assumptions. Consequently, Pickering & Barry's evidence didn't decide between the two kinds of theories, arguably because there are too many ways of conceiving of theories with and without empty elements mediated by movement. At around the same time, Nicol *et al.* (1994) attempted to provide evidence in support of empty element theories using a novel paradigm from the time, cross-modal lexical priming. But McKoon *et al.* (1996) showed that Nicol *et al.*'s evidence was confounded with plausibility, with the consequence that their evidence did not bear on such theories. There were several other papers at the time (e.g. Bever & McElree 1988; MacDonald 1989), but none provided convincing evidence either for or against the hypothesis that empty elements mediate long-distance dependencies.

Thus, so far, no conclusive quantitative evidence favoring a movement analysis over another kind of analysis for a given syntactic phenomenon has been provided. It appears that at this point there are simply too many parameters in current syntactic and semantic theories to provide evidence that would be convincing of either position. We have shown here that in this domain too, HKV's evidence is explained by other linguistic and non-linguistic properties of their materials and hence has no bearing on the question of whether there exist empty elements (in this case, as a result of QR) in linguistic representations. The forces of pragmatics that we argue for are more generally supported by the empirical literature documenting the

power of such pressures, as well as formal and quantitative evidence that such inferences result from rational reasoning about communication (e.g. Frank & Goodman 2012; Bergen *et al.* 2012; Gibson *et al.* 2013).

One possible way for proponents of movement to remedy the empirical situation would be to spell out a specific theory of sentence processing that makes predictions for difficulty arising for the processor at different parts of any given sentence. Lacking such an explicit theory, the most parsimonious assumption is that empty categories, if they exist, do not influence sentence processing in an empirically detectable way.

Acknowledgements

We thank Leon Bergen, Veena Dwivedi, Bart Geurts, Jeremy Hartman, Laura Kertz, David Pesetsky, Geoffrey Pullum, Norvin Richards and two anonymous reviewers for comments on earlier versions of this work and issues related to it. We also thank Jorie Koster-Hale for help in constructing some of the materials in Experiment 5. P.J.'s research was supported in part by NSF Grant BCS 0646081.

Supplementary Data

Supplementary data are available at *Journal of Semantics* online.

EDWARD GIBSON

*Department of Brain and Cognitive Sciences
MIT*

e-mail: egibson@mit.edu

PAULINE JACOBSON

*Department of Cognitive, Linguistic, and
Psychological Sciences*

Brown University

e-mail: pauline_jacobson@brown.edu

PETER GRAFF

Intel Corporation Santa Clara, CA 95054

e-mail: peter.graff@intel.com

KYLE MAHOWALD

Department of Brain and Cognitive Sciences

MIT

e-mail: kylemaho@mit.edu

EVELINA FEDORENKO

*Department of Brain and Cognitive Sciences
MIT*

e-mail: evelina9@mit.edu

STEVEN T. PIANTADOSI

*Department of Brain and Cognitive Sciences,
University of Rochester*

e-mail: spiantadosi@bcs.rochester.edu

REFERENCES

- Amsili, Pascal, Grégoire Winterstein & Emilia Ellsiepen. (2012), 'How obligatory is too? Workshop on Anaphora, University of Paris, Diderot October, 2012.
- Bach, Emmon. (1968), 'Nouns and noun phrases'. In Emmon Bach & Robert Harms (eds.), *Universals in Linguistic Theory*. Holt, Rinehart and Winston. New York. 90–122.
- Barker, Chris. (2002), 'Continuations and the nature of quantification'. *Natural Language Semantics* **10**:211–242.
- Barr, D.J., R. Levy, C. Scheepers & H. Tily. (2013), 'Random-effects structure for confirmatory hypothesis testing: Keep it maximal'. *Journal of Memory and Language* **68**:255–278.
- Bates, Douglas, Martin Maechler & Bin Dai. (2008), lme4: Linear mixed-effects models using S4 classes. R package version 0.999375–27. <http://lme4.r-forge.r-project.org/>.
- Bergen, Leon L., Noah D. Goodman & Roger Levy. (2012), 'That's what she (could have) said: How alternative utterances affect language use'. In Naomi Miyaki, David Peebles & Richard P. Cooper (eds.), *Proceedings of the Thirty-Fourth Annual Conference of the Cognitive Science Society*. Cognitive Science Society. Austin, TX. 120–125.
- Bever, Thomas G & Brian McElree. (1988), 'Empty categories access their antecedents during comprehension'. *Linguistic Inquiry* **19**:35–45.
- Bouton, Laurence F. (1970), 'Antecedent contained pro-forms' In *Papers from the Sixth Regional Meeting of the Chicago Linguistics Society*. Chicago Linguistics Society. Chicago. 154–167.
- Cornack, Annabel. (1984), 'VP anaphora: variables and scope'. In Fred Landman & Frank Veltman (eds.), *Varieties of Formal Semantics*. Foris Publications. Dordrecht. 81–102.
- Evans, Frederic. (1988), 'Binding into anaphoric verb phrases'. In *Proceedings of the 5th Annual Eastern States Conference on Linguistics*. ESCOL Publications Committee, Linguistics Department, Ohio State University. Columbus. 122–129.
- Fodor, Janet Dean. (1998), 'Learning to parse?' *Journal of Psycholinguistic Research* **27**:285–319.
- Frank, Michael & Noah D. Goodman. (2012), 'Predicting pragmatic reasoning in language games'. *Science* **25**: 998.
- Gibson, Edward, Leon Bergen, Steven T & Piantadosi. (2013), 'The rational integration of noisy evidence and prior semantic expectations in sentence interpretation. Proceedings of the National Academy of Sciences'. *Proceedings of the National Academy of Sciences* **110**(20):8051–8056.
- Gibson, Edward & Greg Hickok. (1993), 'Sentence processing with empty categories'. *Language and Cognitive Processes* **8**:147–161.
- Gibson, Edward, Steven T. Piantadosi & Kristina Fedorenko. (2011), 'Using mechanical turk to obtain and analyze English acceptability judgments'. *Language and Linguistics Compass* **5**: 509–524.
- Gibson, Edward, Kyle Mahowald, Roger Levy & Steven T. Piantadosi. (submitted), 'Erroneous analyses drive the reported reading time effects in Hackl, Koster-Hale & Varvoutis (2012)'.
- Hackl, Martin, Jorie Koster-Hale & Jason Varvoutis. (2012), 'Quantification and ACD: evidence from real time sentence processing'. *Journal of Semantics* **29**: 145–206.

- Heim, Irene. (1991), 'Artikel und Definitheit'. *Semantik: Ein Internationales Handbuch der Zeitgenössischen Forschung* 487–535.
- Heim, Irene & Angelika Kratzer. (1998), *Semantics in Generative Grammar*. Blackwell. Malden, MA.
- Hendriks, Herman. (1993), Studied flexibility. Doctoral Dissertation, ILLC.
- Jacobson, Pauline. (1992a), 'Antecedent contained deletion in a variable-free semantics'. In Chris Barker & David Dowty (eds.), *SALT* 2. 193–213.
- Jacobson, Pauline. (1992b), 'Flexible categorial grammars: questions and prospects'. In Robert Levine (ed.), *Formal Grammar: Theory and Implementation*. Oxford University Press. Oxford. 129–167.
- Jacobson, Pauline. (2008), 'Direct compositionality and variable-free semantics: the case of Antecedent Contained Deletion'. In Kyle Johnson (ed.), *Topics in Ellipsis*. Cambridge University Press. Cambridge, UK. 30–68.
- Jacobson, Pauline. (2009), 'Do representations matter or do meanings matter? The case of antecedent containment'. In Erhard Hinrichs & John Nerbonne (eds.), *Theory and Evidence in Semantics*. CSLI publications. Stanford. 81–107.
- Jacobson, Pauline & Edward Gibson. (2014), 'On Line Processing of ACD gives no evidence for Quantifier Raising'. In *Proceedings of Semantics and Linguistic Theory 24*. Cornell Linguistics Club. Ithaca.
- Kaplan, Jeff. (1984), 'Obligatory too in English'. *Language* 60:510–518.
- Kempson, Ruth, Wilfried Meyer-Viol & Dov Gabbay. (2001), *Dynamic Syntax. The Flow of Language Understanding*. Blackwell Publishers, Oxford, UK.
- Lakoff, George. (1971), 'On generative semantics'. In Danny Steinberg & Leon Jakobovits (eds.), *Semantics*. Cambridge University Press. Cambridge. 232–296.
- MacDonald, Maryellen C. (1989), 'Priming effects from gaps to antecedents'. *Language and Cognitive Processes* 4:35–56.
- McCawley, James. (1970), 'Where do noun phrases come from?' In Rodney Jacobs & Peter Rosenbaum (eds.), *Readings in English Transformational Grammar*. Ginn and Company. Waltham. 166–183.
- May, Robert. (1977), 'The grammar of quantification'. Doctoral Dissertation, Massachusetts Institute of Technology.
- Montague, Richard. (1974), 'The proper treatment of quantification in ordinary English'. In Richmond H. Thomason (ed.), *Formal philosophy: Selected papers of Richard Montague*. Yale University Press. New Haven. 247–270.
- McKoon, Gail, Roger Ratcliff & David Albritton. (1996), 'Sentential context effects on lexical decisions with a cross-modal instead of an all-visual procedure'. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22:1494–1497.
- Nicol, Janet L., Janet Dean Fodor & David Swinney. (1994), 'Using cross-modal lexical decision tasks to investigate sentence processing'. *Journal of Experimental Psychology: Learning, Memory and Cognition* 20:1229–1238.
- Partee, Barbara & Mats Rooth. (1983), 'Generalized conjunctio and type ambiguity'. In Rainer Bäuerle, Christoph Schwarze & Arnim von Stechow (eds.), *Meaning, Use and Interpretation of Language*. De Gruyter. 262–283.
- Pickering, Martin & Guy Barry. (1991), 'Sentence processing without empty categories'. *Language and Cognitive Processes* 6:229–259.

- R Core Development Team. (2008), R: a language and environment for statistical computing. Version 2.70. R Foundation for Statistical Computing, Vienna. Available via <http://cran.R-project.org>.
- Rooth, Mats. (1992), 'Ellipsis redundancy and reduction redundancy'. In Steven Berman & Arild Hestvik (eds.), *Proceedings of the Stuttgart Ellipsis Workshop*. Stuttgart.
- Sag, Ivan. (1976), Deletion and logical form. Doctoral Dissertation, Massachusetts Institute of Technology.
- Steedman, Mark. (1987), 'Combinatory grammars and parasitic gaps'. *Natural Language and Linguistic Theory* **5**: 403–439.
- Szabolcsi, Anna. (2013), 'Quantification and ACD: What is the evidence from real time processing evidence for?' *Journal of Semantics* **31**:135–145.
- Tancredi, Christopher D. (1992), Deletion, deaccenting and presupposition. *Doctoral dissertation*, Department of Linguistics and Philosophy, MIT, Cambridge, MA.

First version received: 22.07.2012

Second version received: 29.05.2014

Accepted: 05.06.2014