# A new syntax for multiple *wh*-questions: Evidence from real time sentence processing[*]

Hadas Kotek[1] & Martin Hackl[2],

[1]McGill University & [2]Massachusetts Institute of Technology

Current theories of interrogative syntax/semantics adopt two strategies for the interpretation of in-situ *wh*-phrases: *covert movement* and *in-situ computation*. The covert movement strategy is traditionally assumed to be all-or-nothing: the in-situ *wh*-phrase covertly moves to C or else stays in-situ and is interpreted in its base-generated position at LF. This paper investigates predictions made by these two approaches for real-time processing of English multiple *wh*-questions. We argue that surface in-situ *wh*-phrases are not interpreted in their base-generated positions, but instead undergo covert *wh*-movement. However, covert movement in a question is not unbounded, unlike in current theories that utilize this strategy for the interpretation of a question. Instead, we show evidence that "in-situ" *wh*-phrases undergo a short covert movement step, parallel to the behavior of traditional quantifiers, immediately upon integration into the structure. Following that step, the *wh*-phrase *can* but *need not* move any further: it can be interpreted in its landing site without any further movement. We propose the *partial movement* approach to *wh*-in-situ: A *wh*-phrase can be interpreted at any position with propositional type at LF.

**Keywords:** multiple *wh*-questions, covert movement, in-situ *wh*-phrases, intervention effects, sentence processing

## 1 Introduction

In a language like English, the formation of a question involves at least two steps. First, a structure is formed in which a *wh-phrase* is produced in an argument position of a verb, (1a). Second, that *wh*-phrase is *overtly moved* to the left edge of the sentence, (1b). In a *multiple question* only one *wh*-phrase is pronounced at the left edge of the sentence, while the remaining *wh*-phrase(s) are pronounced in-situ, in, what appears to be, their base-generated positions, (1c).

(1) **Single and multiple questions**
  a. Fred introduced which student to Mary?

[*]For helpful comments, discussions and support, we would like to thank Danny Fox, Irene Heim, Shigeru Miyagawa, David Pesetsky, Michael Yoshitaka Erlewine, Yasutada Sudo, Micha Breakstone, Alexandre Cremers, Alex Drummond, Leo Rosenstein, Michael Wagner, and the audiences of NELS 44, the Amsterdam Colloquium, LSA 2014, and MIT Ling-Lunch. This material is based upon work supported by the National Science Foundation under Grant No. 1251717.

b. *Which student* did Fred introduce __ to Mary?
c. *Which student* did Fred introduce __ to *which professor*?

In the literature on the syntax and semantics of *wh*-questions, there are two schools of thought about the analysis of in-situ *wh*-phrases. Both approaches assume that *wh*-phrases are interpreted by an interrogative head C occurring at the left edge of the question at LF, but they differ in the mechanisms by which C is assumed to create interrogative meanings. Under the first of these approaches, all *wh*-phrases must be structurally adjacent to the C head that interprets them. This approach thus predicts covert movement of in-situ *wh*-phrases as a prerequisite for their interpretation. The second approach invokes mechanisms that interpret *wh*-phrases without any movement. Although these approaches differ substantially in their treatment of interrogative syntax and semantics, traditional means of investigation have not been successful in choosing a preferred account. In particular, these approaches differ in (a) how much covert movement they predict in the derivation of a multiple question, (b) what they assume could cause covert movement, and (c) what syntactic position is targeted by covert movement.

The goal of this paper is to investigate the possible position(s) of in-situ *wh*-phrases at LF. We will argue that both current approaches to *wh*-in-situ are insufficient to derive the correct syntax-semantics of *wh*-questions. We present the results of three self-paced reading experiments that show that the in-situ *wh* in English multiple questions minimally requires a short movement step, comparable to the behavior of quantifiers like *every*, but that from that point on *wh*-phrases can but need not move any further. These results are unpredicted by both approaches to in-situ *wh*-phrases, motivating a new proposal for interrogative syntax/semantics. Specifically, we argue that the results can be explained by a *partial movement* approach to *wh*-in-situ, but not under the traditional approaches.

The remainder of the paper is structured as follows. Section 2 provides background on the two traditional approaches to the interpretation of *wh*-in-situ and the predictions these theories make regarding the source and extent of covert *wh*-movement in a multiple question; it additionally presents the phenomenon of *Antecedent Contained Deletion* (ACD), which has been argued to require covert movement in its derivation, and the methodology that will be used in the experiments in this paper. Section 3 presents two real-time sentence processing studies that test the differing predictions made by the two approaches to *wh*-in-situ for the processing of ACD. We will argue that these experiments lead to two findings: (a) in-situ *wh*-phrases cannot be interpreted in their base position, but (b) they do not necessarily move to C for interpretation. We propose the *partial movement* approach to *wh*-in-situ and show that it makes correct predictions about the processing of multiple *wh*-questions. Section 4 provides additional support for our proposal. We test the prediction that partial movement can be modulated using the phenomenon of intervention effects, and show that this is indeed the case. Section 5 offers a general discussion of the experimental findings and the architecture of the theory that they entail, and a discussion of one alternative interpretation of the results. Section 6 concludes the paper.

# 2 Theoretical background

## 2.1 Two approaches to interrogative syntax/semantics

The literature on the syntax/semantics of *wh*-questions provides two approaches to the interpretation of in-situ *wh*-phrases: covert movement and in-situ interpretation. Under the covert movement approach, all *wh*-phrases must be structurally adjacent to the head that interprets them in the CP periphery. This approach thus predicts covert movement of in-situ *wh*-phrases as a prerequisite for their interpretation. The in-situ approach invokes a mechanism that interprets *wh*-phrases without any movement. Hence, if covert movement does occur in a given question, it must be triggered by other factors. Below is a brief description of the two approaches.

### 2.1.1 The covert movement approach

Under the covert movement approach to questions, *wh*-phrases must be adjacent to C in order to be able to make their contribution to the meaning of the question. Consequently, no *wh*-phrase may remain in situ at LF; instead, all *wh*-phrases occur syntactically next to the complementizer, regardless of where they are pronounced (cf. Karttunen 1977; Huang 1982; Nishigauchi 1986; Lasnik and Saito 1992; Hornstein 1995; Hagstrom 1998; Pesetsky 2000; Richards 2001; Cable 2007, 2010).[1,2]

(2) **The covert movement approach to *wh*-in-situ**
*Which student* [*which professor* [C [Fred introduced __ to __ ]]]

The covert movement approach thus predicts pervasive covert movement in multiple *wh*-questions. Moreover, movement is always triggered for one and the same reason – the semantic needs of the *wh*-phrases themselves – and it always targets the same syntactic position at LF: C. This approach makes the prediction in (3).

(3) **A prediction of the covert movement approach**
All *wh*-phrases in a question must (overtly or covertly) move to C for interpretation.

[1]Other theories argue that at least in some cases, it is not full *wh*-phrases but rather an operator that moves at LF (Aoun and Li, 1993; Hagstrom, 1998; Pesetsky, 2000; Watanabe, 2001; Kishimoto, 2005, a.o.). Since the target position of movement and the reasons for this movement are the same as in the approach introduced here, we classify these theories here as consistent with the covert movement approach. However, as we will see, even if these theories are classified as in-situ theories, they will be unable to derive the experimental results we present in section 3.

[2]Here and throughout, straight arrows indicate movement and curly arrows to indicate an area in which in-situ composition is used. Dashed arrows indicate covert movement. These arrows are used here as a notational convenience only.

### 2.1.2 The in-situ approach

Under the in-situ approach to questions, no (overt or covert) movement is required in order to assign interrogative meaning to a structure containing *wh*-elements (cf. Hamblin, 1973; É Kiss, 1986; Cheng, 1991; Tsai, 1994; Chomsky, 1995; Reinhart, 1998; Kratzer and Shimoyama, 2002; Shimoyama, 2006). The meaning of a question like (4) can be calculated through a mechanism that passes the meanings of *wh*-words up the structure until they reach C, where they can be interpreted (Rooth 1985, 1992). From this perspective, there is no reason to expect any instances of *wh*-movement that are caused by the semantic needs of the *wh*-words themselves. Even the fact that English questions require overt fronting of one *wh*-phrase is unexpected. To explain this fact, a purely syntactic mechanism must be invoked, unrelated to interrogative semantics, for example an 'EPP' feature requiring C to have a filled specifier (Chomsky, 1981).

(4) **The in-situ approach to *wh*-in-situ**
*Which student* [ $C_{+EPP}$ [ Fred introduced __ to *which professor* ]]

The in-situ approach to questions thus makes no predictions about the position of *wh*-phrases at LF. Following standard assumptions in the theoretical literature that the simplest syntactic structure for a sentence is always preferred to a less simple one (cf. Chomsky, 1991, 1993, 1995, 2000; Epstein, 1992; Kitahara, 1997; Fox, 2000; Collins, 2001; Richards, 2001; Reinhart, 2006), it is predicted that *wh*-phrases occupy the position at which they were merged into the syntactic structure, unless another syntactic process triggers their movement. No covert movement occurs for semantic interpretation of *wh*-phrases, (5).

(5) **A prediction of the in-situ approach**
*Wh*-phrases in a question can be interpreted in situ and do not require any movement.

## 2.2 Antecedent Contained Deletion

In this section we discuss Antecedent Contained Deletion (ACD), which will be central to the experimental methodology introduced in section 2.3 and used in sections 3-4.

ACD is a phenomenon found in certain VP ellipsis contexts. For VP-ellipsis to be licensed, a pronounced antecedent VP must exist that is identical to the missing VP.[3] This is straightforward in examples like (6), where the only pronounced VP in the sentence—*read a book*—can serve as an antecedent for the missing VP. In ACD cases like (7), however, the missing

[3] How to define identity in the domain of ellipsis is a much debated question. For example, experiments have shown that voice mismatches are possible in some contexts but not others (e.g. Kehler, 2001; Arregui et al., 2006; Kertz, 2010; San Pietro et al., 2012, a.o.). It has also been argued that such mismatches are only possible in VP ellipsis but not in sluicing (Merchant, 2013). We will not attempt to contribute to the definition of identity here. We assume that traces count as identical for the purpose of ellipsis parallelism if they bound from parallel positions (e.g. Fox, 2002).

VP appears to be properly contained inside the only possible antecedent VP in the sentence. Matching the missing VP with the antecedent VP should be impossible since the missing VP is properly contained inside its antecedent and so cannot be identical to it.

(6) **VP-ellipsis and its resolution**

a. John read a book and Mary did ⟨missing VP⟩, too.

b. John read a book and Mary did ~~⟨read a book⟩~~, too.

(7) **ACD and the containment problem**

a. John read every book that Mary did ⟨missing VP⟩.

b. John read every book that Mary did ⟨??⟩.

To solve the containment problem in (8a), the standard analysis of ACD assumes that the object is covertly moved out of $TP_1$ to a syntactic position in the higher $TP_2$, yielding the LF in (8b). The resulting VP, containing only the Verb and the trace of covert movement, can then be used as an antecedent for the missing VP (cf. Williams, 1974, 1977; Sag, 1976; May, 1985; Larson and May, 1990; Johnson, 1996; Heim and Kratzer, 1998; Fox, 2003).

(8) **Resolution of ACD using covert movement**

a. [$_{TP1}$ John read [every book that Mary did ⟨missing VP⟩ ].

b. [$_{TP2}$ [every book that Mary did ~~read t~~ ] [$_{TP1}$ John read t ]].

ACD can also occur in sentences with a *wh*-phrase complement of the verb. Example (9a) illustrates this for a sentence with an embedded multiple question, where the in-situ *wh*-phrase hosts an ACD site. The containment problem for this sentence is illustrated in (9b-c). To undo containment and allow for ACD resolution, the in-situ *wh*-phrase must undergo covert movement. Once covert movement has occurred, an appropriate antecedent can be found and used for the missing VP, (9d).

(9) **ACD in multiple *wh*-questions and its resolution**

a. John knows [$_{TP1}$ which student read [which book that Mary also did ⟨missing VP⟩ ]].

b. John knows [$_{TP1}$ which student read [which book that Mary also did ~~⟨read [which book that Mary also did ⟨missing VP⟩⟩~~ ]]].

c. John knows [$_{TP1}$ which student read [which book that Mary also did ~~⟨read [which book that Mary also did read [which book that Mary also did ⟨missing VP⟩⟩~~]]]].

d. John knows [$_{TP1}$ which student [$_{TP2}$ which book that Mary also did ~~⟨read t⟩~~ ] [ read t ]]

## 2.3 Self-paced reading and the Hackl et al. (2012) paradigm

In this section we present the methodology that will be used in the experiments in section 3, whose goal is to investigate the presence and extent of covert *wh*-movement in multiple questions. These experiments will use the paradigm developed in Hackl et al. (2012), which studies the real-time processing of sentences where covert movement has been argued to occur, concentrating on the universal quantifier *every*.

This paradigm takes advantage of the inherently linear organization of real-time sentence processing: the human parser integrates material that occurs earlier in a sentence before it encounters material occurring later. The paradigm relies on two underlying assumptions about the economy of structure building: (a) the linguistic parser always builds the simplest syntactic structure consistent with the linguistic input (cf. Bever, 1970; Frazier and Rayner, 1982; Phillips, 2003), and (b) structures without covert movement are simpler than structures with covert movement (cf. Fox, 1995, 2000; Tunstall, 1998; Frazier, 1999; Anderson, 2004). Given these assumptions, we expect the parser not to postulate covert movement in the parse of a sentence until the point at which the parser determines that it is necessary. From that point on, one might expect to find online consequences (detectable as a delay in Reading Times, RTs) of the *reanalysis* of the structure to a less preferred parse.

Hackl et al. (2012) compare the processing of sentences that contain non-quantificational objects (*the*, 10a-c) to sentences with quantificational objects (*every*, 10d-f)—where this factor is crossed with three different gap sizes of ACD inside an attached relative clause: (a) *no ellipsis* using a lexical verb, (b) *small ellipsis* marked by did, where the antecedent of the ACD is the embedded $VP_2$ headed by *treat*, and (c) *large ellipsis* marked by was, where the antecedent of the ACD is the matrix $VP_1$ headed by *reluctant*.

(10) **The paradigm in Hackl et al. (2012)**

The doctor was [$_{VP1}$ reluctant to [$_{VP2}$ treat ...

a. **the** patient that the recently hired nurse *admitted* (*no ellipsis*)
b. **the** patient that the recently hired nurse *did* (*small ellipsis*)
c. **the** patient that the recently hired nurse *was* (*large ellipsis*)
d. **every** patient that the recently hired nurse *admitted* (*no ellipsis*)
e. **every** patient that the recently hired nurse *did* (*small ellipsis*)
f. **every** patient that the recently hired nurse *was* (*large ellipsis*)

... after looking over the test results.

ACD resolution involves at least three steps: (i) creating a structure in which antecedent containment is undone; (ii) identifying an appropriate antecedent for the ellipsis; and (iii) filling the antecedent into the gap and computation of the resulting meaning. Steps (ii) and (iii) are required for all cases of VP ellipsis. Step (i)—the reanalysis of the structure so as to undo antecedent containment—is only required in the case of ACD. Hackl et al. (2012) generate specific predictions for language processing in real time based on these properties and the assumptions (a) that step (i) of ACD resolution requires covert movement (cf. section 2.2), and (b) that quantificational objects (but not non-quantificational objects) require covert movement for their interpretation, and that this movement targets the lowest position in the structure where the object can be interpreted (cf. May, 1985; Fox, 1995, 2000; Heim and Kratzer, 1998)

In the *definite* conditions in (10a-c), no covert movement is predicted to take place when the definite article is processed. The parser only assumes covert movement after it has been determined that the sentence contains an instance of ACD. This happens after encountering the auxiliaries <u>*did*</u> and <u>*was*</u> in (10b-c), which, together with the immediately following word, signal the presence of an ACD site and thus trigger reanalysis in order to resolve ACD. As discussed above, this reanalysis should incur some processing cost detectable in RTs following the ellipsis site, compared to the baseline with *no ellipsis* (10a). Furthermore, the difference in the locality of covert movement and, concomitantly, the size of the antecedent VP should also be reflected in RTs: the covert movement in (10c) must target a non-local position, above the matrix $VP_1$ headed by *reluctant*, in order to make the matrix VP available for ACD resolution while the covert movement in (10b) targets a closer position above the embedded $VP_2$ headed by *treat*. On the assumption that non-local movement and the retrieval of a larger antecedent VP are more costly than local movement and the retrieval of a smaller antecedent VP, (10c) is expected to produce longer RTs than (10b). These predictions are summarized in (11). Here and below, triggers for movement are boxed:

(11) **Predictions for the *definite* conditions**

a. The doctor was reluctant to treat **the** patient that the nurse *admitted.*
   ⇒ Lexical verb condition: No movement assumed at all.

b. The doctor was reluctant to treat **the** patient that the nurse $\boxed{\textit{did.}}$
   ⇒ Movement assumed after **did** is encountered. *late trigger*
   Movement targets a position above *treat*. *short movement*

c. The doctor was reluctant to treat **the** patient that the nurse $\boxed{\textit{was.}}$
   ⇒ Movement assumed after **was** is encountered. *late trigger*
   Movement targets a position above *reluctant*. *long movement*

For the *quantificational* conditions, different predictions are made. In all of (10d-f), the parser must assume covert movement as soon as it encounters the quantificational object headed by *every*. Furthermore, the movement is expected to be *local* and to target a position just above the embedded $VP_2$ headed by *treat* and no further: this is the first position where quantifiers in object position can be integrated into the structure without causing a type-mismatch. Importantly, this position is high enough to preemptively undo antecedent containment in the case of *small ellipsis* (10e) but not in the case of the *large ellipsis* (10f): the movement triggered by *every* targets the same position that would be independently targeted for the resolution of the *small ellipsis*. Because this movement has already happened earlier in the parse, step (i) of ACD resolution can be avoided, and only steps (ii)-(iii) must apply when the ellipsis marker *did* is encountered. This means that at the point of identifying the ACD site no *reanalysis* is predicted for this sentence. Hence, ACD resolution is expected to be easier compared to (10b) since part of the work necessary to resolve ACD has already happened prior to encountering the ACD marker *did* with *every*, but all three steps of ACD resolution must happen when *did* is encountered with *the*.

In the case of the *large ellipsis* in (10f), by contrast, the covert movement step that was assumed following the processing of *every* is not sufficient to furnish a suitable antecedent VP: following this local movement, the missing VP is still contained inside its antecedent. Hence, when the auxiliary *was* is reached, the parser must again reanalyze the structure, covertly moving the object a second time, from its position above the embedded VP to a position higher than the matrix VP. This means that no facilitation of ACD resolution is expected in (10f) even though the host DP is quantificational in nature. These predictions are summarized in (12):

(12) **Predictions for the *every* conditions**

a. The doctor was reluctant to treat **every** patient that the nurse *admitted.*
⇒ Movement assumed after **every** is encountered. — *early trigger*
Movement targets a position above *treat*. — *short movement*

b. The doctor was reluctant to treat **every** patient that the nurse *did.*
⇒ Movement assumed after **every** is encountered. — *early trigger*
Movement targets a position above *treat*. — *short movement*

c. The doctor was reluctant to treat **every** patient that the nurse *was.*
⇒ Movement assumed after **every** is encountered. — *early trigger*
Movement targets a position above *treat*. — *short movement*
⇒ **Additional** movement assumed after **was** is encountered. — *late trigger*
Movement targets a position above *reluctant*. — *long movement*

Hackl et al. (2012) used sentences in a paradigm as in (10) in a self-paced reading study (Just et al., 1982): participants read sentences that appeared on the screen one word at a time using a moving window display. Residual Reading Times (RRTs) were analyzed for each word in the sentence.

Figure 1 shows RRTs two words after the ellipsis site. These results indicate that the predictions described above are borne out: When the object is *definite*, ACD resolution is associated with longer RRTs compared to the baseline condition, *no ellipsis* (*Verb*). The increase in RRTs is linear across the three gap size levels (*the-verb* vs. *the-did* vs. *the-was*). When the object is *quantificational*, however, an interaction pattern emerges. No increase is observed between the *every-verb* condition and the *every-did* condition (*small ellipsis*), while a large increase is observed between the *every-did* condition and the *every-was* condition (*large ellipsis*). This is unexpected if the two factors (quantificational vs. non-quantificational object and ACD size) are not linked in some form, but it is the expected result under the assumption that covert movement is required to accommodate both a quantificational object and an ACD site, as the first step in resolving antecedent-contained ellipsis.

There is a debate in the sentence processing literature about whether or not the size of the antecedent site affects the processing of the ellipsis site (cf. Murphy, 1985; Frazier and Clifton, 2000, 2001; Martin and McElree, 2008, 2009, 2011). We note that the results discussed here do not rely on the answer to this question. Although the current thinking in the literature seems to suggest that a larger antecedent does not necessarily incur a larger cost in the processing of ellipsis, it is possible to attribute the increased processing cost that Hackl et al. (2012) observe for the non-local ACD cases either to the longer covert movement step associated with the *large ellipsis* condition, or to the fact that the *large ellipsis* involves a more complex VP and hence more complex meaning than the *small ellipsis*. Another possible alternative proposed by an anonymous reviewer is that the parser's search for the antecedent is bounded by the structure. That is, regardless of the type of auxiliary, the parser attempts to resolve the ellipsis to the local potential antecedent. Here, the parser first attempts to resolve the ellipsis to the local VP, which is a candidate as long as the object DP that hosts the ACD site has been QRed above that VP. However, in the case of *large ellipsis* the auxiliaries do not match, and this mismatch triggers a further search which, in turn, involves re-movement of the object DP—via a second QR step—and re-resolution of the ellipsis site. These additional operations incur a greater processing cost.

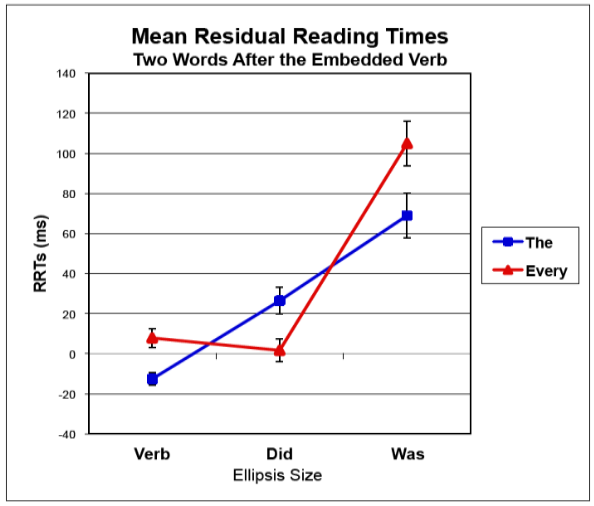



Figure 1: Results of Hackl et al. (2012)

## 2.4 How syntax and online processing interact

At this point, let us step back and more carefully spell out the assumptions that self-paced reading paradigms rely on. This discussion will serve as the basis for the experiments presented in section 3 below, whose purpose is to detect the LF position of surface in-situ *wh*-phrases in English multiple questions, using the Hackl et al. (2012) paradigm.

### 2.4.1 Structure building and processing

We begin by discussing how online sentence processing corresponds to structure building in syntax. The Self-Paced Reading paradigm assumes that linguistic structure can be built in real time from left to right. This is enforced rather straightforwardly by the structure of the paradigm: participants encounter material on the left before they reach material on the right, and hence processing of linguistic material proceeds similarly from left to right (Phillips, 1996).

Next, we may wonder when the parser begins to build a parse for the material it encounters. A consensus in the literature is that the parser does not wait until a full sentence has been uttered in order to begin imposing a structure on this sentence. Instead, structure is built 'online,' in tandem with reading or hearing the sentence. The literature offers two approaches to this process: *serial processing*, where the parser pursues just one parse of a sentence at a time (e.g. Ford et al., 1982; Frazier and Rayner, 1982; Pritchett, 1992) and *parallel processing*, where the parser pursues multiple analyses at once (e.g. MacDonald et al., 1994; Tanenhaus and Trueswell, 1995). An intermediate third way of viewing processing is to assume a *serial* parser, in the sense that it assumes one particular parse of the sentence at a time, but at each step multiple potential parses are evaluated before this step is made. Here we will be using rhetoric that assumes a serial parse, but the conclusions we draw are compatible with parallel models as well. The important assumption—a standard and well-motivated one—is that the parse is built simultaneously with reading or hearing the sentence, not after it has been completed.

Although processing happens from left-to-right and is very fast, it may not be instantaneous or strictly incremental in the sense that each new word is fully processed and integrated before the next word is processed. In fact, effects in the self-paced reading paradigm often show up not on the word expected to trigger an effect, but several words downstream.[4] The more costly the expected computation, the longer it takes for the effect to show up. For this reason, it is advisable to employ precaution measures that make it possible to isolate the hypothesized effects of specific triggers in a sentence. First, possible triggers for slowdown—for example, suspected triggers of covert movement—are separated from one another by additional lexical material. In the Hackl et al. (2012) paradigm in (10), two factors are varied: the quantifier heading the relative clause (*the* vs. *every*), and the ellipsis marker at the right edge of that clause (*verb* vs. *did* vs. *was*). These two factors are separated by 6 words, to ensure that effects from the earlier trigger do not interfere with the later trigger.[5] Second, a "continuation" is used following the second trigger, to allow for a spillover region where effects of this trigger might show up, if they are not instantaneous. These can sometimes make for a long and perhaps cumbersome sentence, but this complication is an inherent component of all SPR experiments, and experimenters work to ensure that while long, the sentences are as easy to process as possible.

### 2.4.2 Covert movement and Self-Paced Reading

With this in mind, consider how movement is integrated into an online parse of a sentence. Overt movements (to the left) constitute *filler-gap dependencies*, which have a well-known processing signature: the parser encounters a linguistic element in a position other than its theta-position; this element must be stored in memory until its gap position is found, so that it can be integrated into the structure there. There is a rich literature surrounding the

[4] We return to a detailed discussion of this fact in section 2.4.4.

[5] See graphs and discussion in Hackl et al. (2012) that show that this is indeed the case in their experiments. As we will see below, this is also the case in the experiments presented here.

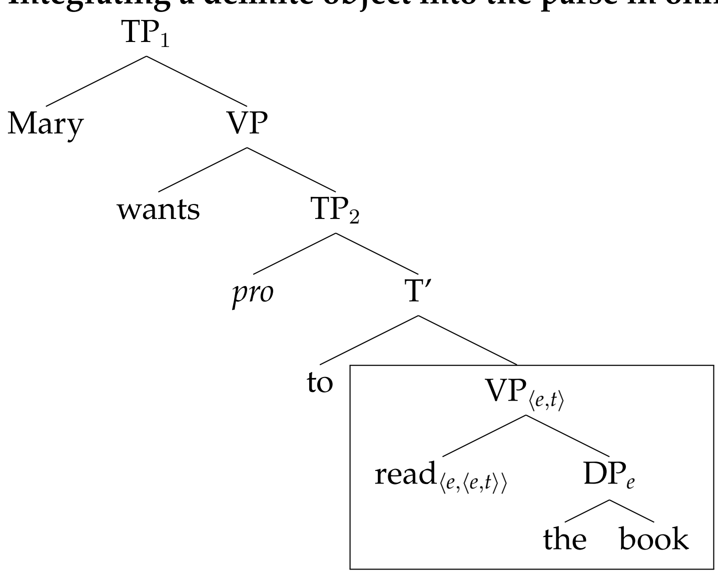
process which is implicated in processing such filler-gap dependencies. For example, it is known that the difficulty of processing these dependencies increases with distance (Gibson, 1998), interference effects (Gordon et al., 2004; van Dyke and McElree, 2006), as well as processing load imposed by other referential entities along the filler-gap path (Warren and Gibson, 2002).

Much less is known about the effects of covert movement operations on real time sentence processing. To get off the ground, we need to make explicit two assumptions that will play a role in our reasoning: First, we assume, following Frazier and Fodor (1978) and much subsequent work, that processing is not just incremental from left-to-right but also biased towards simplicity along the lines of the principle of Minimal Attachment, (13) (see also Bever, 1970; Frazier and Rayner, 1982; Phillips, 2003, a.o.). Second, we assume that, all else being equal, a parse without covert movement is simpler than a parse with covert movement, and that a parse with local covert movement is simpler than a parse with non-local covert movement (cf. Fox, 1995, 2000; Tunstall, 1998; Frazier, 1999; Anderson, 2004).

(13) **The minimal attachment principle**
Attach each new item into the current phrase marker postulating only as many syntactic phrase nodes as is required by the grammar.

Given these assumptions, we expect the parser not to postulate covert movement in the parse of a sentence until the point at which the parser determines that it is necessary. Therefore, we expect the parser to assume no covert movement upon encountering a definite object DP, since definite descriptions are able to semantically merge with a verb in-situ. This step in the parse is illustrated (14).[6]

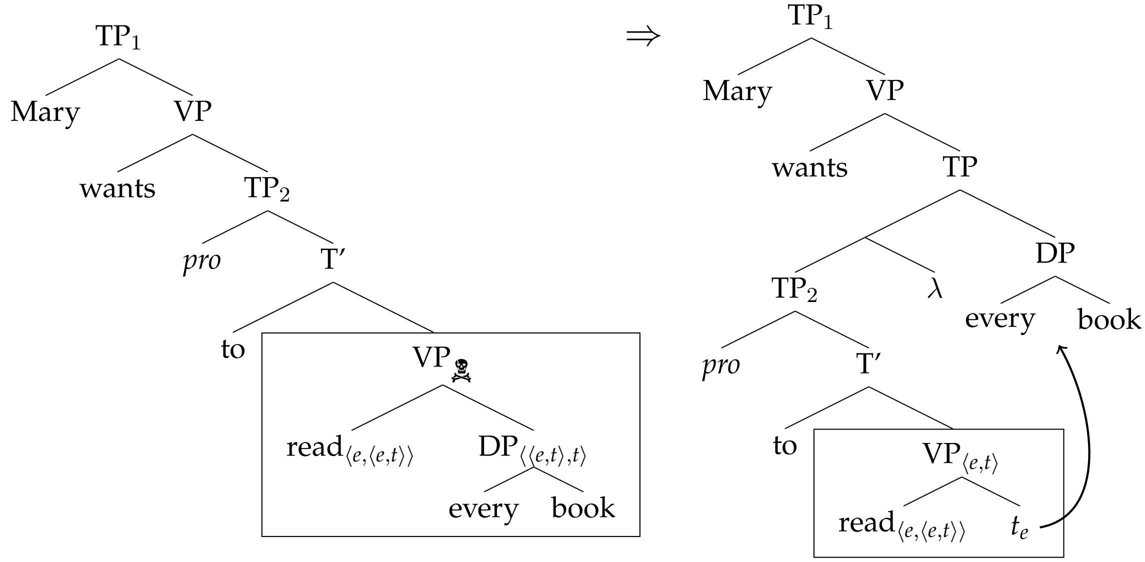
(14) **Integrating a definite object into the parse in online processing**




[6]Extensional types are shown just for the relevant part of the structure. For simplicity, we abstract away from the VP-internal subject hypothesis.

Conversely, we expect movement to be assumed when a quantificational object is encountered. Following the principles laid out above, we expect this movement to be as short as possible, to avoid assuming more structure than necessary. This step in the parse is illustrated (15). The quantificational object cannot be integrated into the structure as the sister of the verb, because of a type mismatch. The object is instead integrated into the first position at which it is interpretable—here, immediately above $TP_2$.

(15) **Integrating a quantificational object into the parse in online processing**




Since the change from the first parse in (15) to the second amounts to a *reanalysis* triggered by the quantificational DP 'every book,' one might expect to find online consequences (detectable as a delay in Reading Times, RTs) of the move to a less preferred parse. Alternatively, a slowdown might be attributed simply to the presence of the quantifier in the structure: since a quantifier is less frequent in text and speech than a definite article, a slowdown in RTs might reflect surprisal due to lexical frequency and nothing more. As a result, an effect of *every* vs. *the* can only be taken to mean that readers have noticed the difference between these two determiners, and nothing further. Moreover, it is worth pointing out that observing no slowdown after a quantificational object does not mean that reanalysis as described in (15) did not take place. It could simply be that this sort of reanalysis is preformed by the parser with sufficient automaticity that SPR is not sensitive enough to detect it. As a result, an effect or the lack of an effect here cannot be conclusively interpreted to be associated with a covert movement step in the derivation.

Finally, note that unlike in cases of overt movement, we do not expect costs associated with a filler-gap dependency: once the quantificational object is integrated into the structure, the parser already knows where the gap position of movement is, and therefore it need not hold onto the filler until a gap is located. That is, once the movement is done, the meaning of the structure can be computed immediately and no further operation is required.

### 2.4.3 Ellipsis resolution and Self-Paced Reading

The Hackl et al. (2012) paradigm crucially relies on a process of ellipsis resolution that is triggered once the auxiliaries *did* or *was* are encountered in the sentence. As discussed in the previous section, ellipsis resolution involves at least two steps: (i) identifying an appropriate antecedent for the ellipsis; and (ii) filling the antecedent into the gap and computing the resulting meaning of the sentence. In the case of Antecedent Contained Deletion, an additional step is necessary: an appropriate antecedent must first be constructed, before steps (i) and (ii) above can proceed. As section 2.2 shows, in the Hackl et al. (2012) paradigm, this involves covert movement of the object to a position above the verb heading the antecedent VP.

The movement step involved in this process may proceed in more than one way, as discussed above: we may assume that the parser always assumes movement above the most local VP, as a first step. The parser then attempts to use this VP as antecedent. If the resulting meaning is convergent, the parse may continue. This is the case when *small ellipsis* (with *did*) is encountered. Otherwise, the parser must assume additional movement, and will attempt to use the next higher VP as the antecedent for ellipsis. In the Hackl et al. (2012) paradigm, this happens when the auxiliary *was* is used, requiring a *large ellipsis* targeting a non-local antecedent VP. Thus, this process is more costly in the case of *large* ellipsis than in the case of *small* ellipsis. Alternatively, we may assume that the parser is able to immediately target the correct antecedent in both the *small* and *large* ellipsis conditions. In this case too, however, the large ellipsis has a richer content than the smaller ellipsis, and therefore integrating it into the structure and computing the resulting meaning of the sentence may be more costly. Notice that this assumption might be made even if we entertain a theory of ACD resolution that does not involve movement.

Under either approach, we may expect to find slower RTs with larger ellipsis than with smaller ellipsis. As we will see below, such an effect is indeed present in all of the experiments presented in section 3. Thus, a main effect of ellipsis size alone will not be taken as indication for movement. Instead, the crucial question will be whether an earlier trigger for movement facilitates ACD resolution downstream. This is expected under a movement analysis of ACD, but is unexpected if ACD resolution does not involve movement.

### 2.4.4 The timing of effects in Self-Paced Reading and their interpretation

Finally, we can now ask when we might expect to detect an effect of covert movement in a Self-Paced Reading paradigm. Obviously, one certain prediction is that we should not see an effect *before* a trigger is encountered. However, as mentioned above, it is common to find an effect not *on* the trigger word itself, but in the spillover region *following* that word (e.g. Clifton et al., 1999; Wagers and Phillips, 2009; Kazanina and Phillips, 2000; Xiang et al., 2011; Polinsky et al., 2012; Polinsky and Potsdam, Forthcoming, among others). Of course, it is doubtful whether effects that occur far downstream for a trigger should be attributed (solely) to that trigger, but it is common for effects to show up in the near vicinity of the trigger. The more costly the computation associated with the trigger, the further downstream an effect may occur.

In the case of the Hackl et al. (2012) paradigm, the crucial region of interest starts with the later trigger for movement—the auxiliary verbs *did* and *was*, which signal the presence of Antecedent Containment in the structure. Note, however, that the earliest point at which the parser has conclusive evidence for the presence of an ACD site is arguably the word following the auxiliary verb and not the auxiliary itself: at this point in the sentence, readers encounter a conjunction or disjunction that can only be attached at the clausal level, indicating that the previous clause contained ellipsis. For ease of exposition, we will assume in what follows that the trigger is the auxiliary verb itself and start counting words in the spillover region from there on. The question, then, is when we should expect to detect effects of covert movement as revealed by facilitation of the processing an ACD site.

As discussed above, the Hackl et al. (2012) effects appears not on the auxiliary verb itself, but two words after the verb. Below we will see that all five experiments presented in this paper also exhibit an effect in this same region: effects consistently show up two and three words after the auxiliary verb, inside the spillover region.

In this context, it is again important to recall this inherent limitation of the SPR paradigm: the SPR paradigm does not allow for an a-priori prediction as to the precise location of an effect, as long as the effect happens inside the spillover region. To interpret the effects of experiments based on the Hackl et al. (2012) paradigm as stemming from covert movement triggered by the need to resolve ACD, the effect must happen only after the parser encounters the triggers *did*/*was*, and they must happen within a region that is uniform across all conditions so that nothing else in that region could have generated the effects. Our materials below are constructed to ensure this. Moreover, an important strategy for confirming the validity of results of SPR experiments is through replication. Below we will show that crucial effects reported in our experiments survive through repetition of the experiments.

# 3 Against an in-situ approach to *wh*-in-situ

## 3.1 Predictions of *wh*-in-situ theories for the Hackl et al. paradigm

The goal of the present paper is to provide experimental evidence to distinguish between the covert movement approach and the in-situ approach to interrogative syntax/semantics. As we have seen, the two approaches make different predictions regarding the availability of covert movement for in-situ *wh*-phrases in multiple questions, and in particular regarding (a) how much covert movement is predicted to be involved in the derivation of a multiple question, (b) what causes covert movement, and (c) what syntactic position is targeted by covert movement. These predictions are summarized in (16).

(16) **Predictions of the two approaches with regard to covert *wh*-movement**

a. The covert movement approach:
All *wh*-phrases in a question must (overtly or covertly) move to C for interpretation.

b. The in-situ approach:
*Wh*-phrases in a question can be interpreted in situ and do not require any movement.

As the title of this section reveals, we will show evidence that the in-situ approach to *wh*-in-situ cannot be correct for English multiple questions. We will furthermore argue for a particular view of the covert movement approach that is different than the traditional one.

Within the Hackl et al. (2012) paradigm, the covert movement approach and the in-situ approach to the interpretation of in-situ *wh*-phrases make different predictions with regard to facilitation of ACD resolution in sentences in which an ACD gap site is hosted inside a *wh*-phrase, e.g. (17a-b):

(17) **ACD hosted by *wh*-phrase in the Hackl et al. paradigm**

The conductor asked *which soloist* was [$_{VP1}$ willing to [$_{VP2}$ perform...

a. **which** concerto that the brilliant protégé *did* ... *(small ellipsis)*

b. **which** concerto that the brilliant protégé *was* ... *(large ellipsis)*

...and restructured the rehearsal accordingly.

In the covert movement approach, all *wh*-phrases must move *non-locally* to C for interpretation. As a result, we predict that both *small* and *large* ellipsis should be relatively easy to process, because antecedent containment will have been undone by covert *wh*-movement before the parser reaches the gap site. When the parser encounters the gap site, all it needs to do is find an antecedent for the missing VP and nothing more.

In the in-situ approach, *wh*-phrases can be interpreted *in-situ* and hence the parser will not move the in-situ *wh*-phrase to C when it is encountered. Upon reaching the ellipsis site, the parser must reanalyze the structure and covertly move the *wh*-object to a position above the missing VP, in order to construct an appropriate antecedent for the ellipsis. Hence, a high processing cost should be associated with the resolution of the ACD site in both the *small ellipsis* condition and in the *large ellipsis* condition. The predictions of the two approaches are summarized in (18), and will be tested in Experiment 1 below.

(18) **Predictions of the two approaches with regard to ACD resolution**

a. The covert movement approach:
The resolution of both small and large ellipsis is facilitated.

b. The in-situ approach:
The resolution of both small and large ellipsis is not facilitated.

In the following sections we present the results of three experiments that test the predictions of the covert movement approach and the in-situ approach to *wh*-in-situ. All three experiments use the paradigm in Hackl et al. (2012) as a tool for detecting covert movement in multiple *wh*- questions.

Previous processing work, as well as rating and corpus studies, show that questions with D-linked *wh*-phrases (Pesetsky, 1987) are easier to process than questions with bare *wh*-pronouns (Arnon et al., To appear; Clifton et al., 2006; Fanselow et al., 2008; Featherston, 2005a,b; Frazier and Clifton, 2002; Hofmeister et al., 2007), and furthermore that *superiority-obeying* questions are easier to process than *superiority-violating* questions. In order to make the experimental items as easy to process as possible, the items in Experiments 1-3 all use D-linked *wh*-phrases in a superiority-obeying word-order.

## 3.2 Experiment 1: *every* vs. *which*

This experiment tests the core predictions of the covert movement approach and in-situ approach to in-situ *wh*-phrases. In particular, we compare the behavior of questions with an ACD gap hosted by a *wh*-phrase with the behavior of questions with an ACD gap hosted by the quantificational determiner *every*.

### 3.2.1 Design and materials

Experiment 1 presented participants with (embedded) *wh*-questions headed by a subject *wh*-phrase. Two factors were crossed: (a) *determiner*: whether the embedded question contained the quantificational determiner *every*, yielding a simplex *wh*-question, or a second *wh*-phrase, yielding a multiple *wh*-question; and (b) *ellipsis size*: whether the sentence contained a *small ellipsis* marked by <u>did</u>, where the antecedent of the ACD site is the embedded VP, $VP_2$, or *large ellipsis* marked by <u>was</u>, where the antecedent of the ACD site is the matrix VP, $VP_1$. A sample item is given in (19) below:[7]

(19) **Sample target item in Experiment 1**

The conductor asked *which soloist* was [$_{VP1}$ willing to [$_{VP2}$ perform...

a. **every** concerto that the brilliant protégé <u>did</u> ... *(small ellipsis)*

b. **which** concerto that the brilliant protégé <u>did</u> ... *(small ellipsis)*

c. **every** concerto that the brilliant protégé <u>was</u> ... *(large ellipsis)*

d. **which** concerto that the brilliant protégé <u>was</u> ... *(large ellipsis)*

...and restructured the rehearsal accordingly.

[7] To simplify the discussion, in what follows we ignore the highest predicate embedding the question. We thus refer to the lower VP inside the embedded question as the "embedded VP" and to the larger VP inside the embedded question as the "matrix VP."

There were 28 sentence templates following the sample paradigm in (19). Each sentence in a template employed either *every* or *which* as the determiner of the object DP. This DP hosted a relative clause with an ACD site marked with an auxiliary verb. The auxiliary *did* marked a *small ellipsis* corresponding to the embedded predicate, $VP_2$, and the auxiliary *was* marked a *large ellipsis* corresponding to the matrix predicate, $VP_1$. After the ellipsis site, the sentences had continuations beginning with a clausal conjunction or disjunction, which varied in length but were at least 5 words long, providing a spillover region for detecting possible processing difficulties associated with ACD resolution. See Appendix A for a full list of the materials, and Appendix B for an offline ratings study focusing on the acceptability of these items.

Because the experiments were conducted online, it was not possible to control the participants' screen size. Consequently, in order to ensure that the region of interest was read without interruptions that may artificially affect the data, all the sentences were presented on two lines, with the line break in target sentences always placed immediately following the verbal complex (that is, the first line of the sentence was the first line in (19), and the second line contained the text in lines a-d and the continuation following these lines).[8] Target items were counterbalanced across four lists using a Latin Square design and combined with 48 filler sentences of various types resulting in a total of 76 sentences. Non-target items included sentences that were similar to the target items in structure, in length and in containing quantifiers. 18 filler sentences resembled the target sentences in all aspects but contained a lexical verb instead of ellipsis (*did*, *was*). Filler sentences additionally contained line breaks in different positions, making it impossible for participants to anticipate where a line break might occur, or whether or not there will be ellipsis in the sentence. The remaining filler sentences were taken from an unrelated study.

Each experimental item was followed by a yes/no comprehension question. The questions asked about different aspects of the sentences, including about material inside the relative clause and about the predicates used in the sentences, to ensure that participants were processing all parts of the sentence at a deep level. The correct answers to half of the questions was *yes* and to the other half *no*.

### 3.2.2 Methods

Experiment 1 used the moving window self-paced reading methodology and was hosted on Ibex Farm.[9] Participants were presented with sentences that appeared on the screen one word at a time in a moving window display. Each trial begins with a series of dashes marking the length of the sentence. Participants press the spacebar to reveal the next word of the sentence. Each press of the spacebar reveals a new word while the previous word is

[8]Testing on a number of standard monitors showed this method to consistently avoid line breaks inside the region of interest, and was furthermore found to be the most natural among several other options in a pilot study.

[9]Ibex: Internet Based Experiments, created and maintained by Alex Drummond, accessible at http://spellout.net/ibexfarm/.

again replaced with dashes. The amount of time a participant spends reading each word is recorded (RT). After the final word of each sentence, a yes/no comprehension question appears, asking about information contained in the sentence. Participants responded by pressing "1" for "Yes" or "2" for "No." No feedback was given about whether the answer to the question was correct or incorrect.

Before beginning the experiment, participants were given detailed instructions about the experiment and then read and accepted a consent statement. Participants were instructed to read the sentences at a natural rate to ensure understanding. They were also instructed to answer the comprehension questions as accurately as possible. There were three practice items before the experiment began. Each experiment took approximately 30 minutes to complete.

Subjects for the experiments were recruited through Amazon Mechanical Turk and they were paid $1.5 for their participation. The participants were asked about their native language but were told that payment was not contingent on their response. To further ensure that only native speakers of English participated in the experiments, IP addresses of participants were restricted to the US using Amazon Mechanical Turk's user interface. Participation was further restricted so that only Turk Workers with a overall approval rate of over 95% of all their submissions were allowed to participate in the experiments.

### 3.2.3 Predictions

The *every* conditions are expected to replicate the results of Hackl et al. (2012). That is, we expect to find a main effect of ellipsis size, such that *small ellipsis* is easier to process than *large ellipsis*: since *every* triggers covert movement to a position above the embedded $VP_2$ (headed by *perform* in (19)) as soon as the quantifier is encountered, we expect antecedent containment to be preemptively undone in the case of the *small ellipsis* (marked with *did*) (19a), leading to facilitation of ACD resolution in this case. However, since this movement does not target a position high enough to undo antecedent containment in the case of the *large ellipsis* (marked with *was*) (19c), we expect ACD resolution to be relatively more difficult in this case: once the auxiliary *was* is reached, the parser must perform a second reanalysis, covertly moving *every* from its QRed position above the embedded $VP_2$ to a position above the higher $VP_1$ (headed by *willing* in (19)), in order to allow for ACD resolution. These two conditions thus provide us with a baseline contrast against which to compare the *which* conditions in (19b,d).

For the *which* conditions, the two approaches to in-situ *wh*-phrases make different predictions:

In the covert movement approach, all *wh*-phrases must move *non-locally* to C for interpretation. Both *small* and *large* ellipsis are predicted to be relatively easy to process because antecedent containment is undone by covert *wh*-movement before the parser reaches the ellipsis site. When the parser encounters the ellipsis site—regardless of whether the ellipsis is *small* or *large*—all that is left to do is find an antecedent for the missing VP. We thus expect an interaction, such that the *which-did* and *which-was* conditions both pattern with the *every-did* condition and exhibit facilitation effects, compared to the remaining *every-was* condition where we predict participants to exhibit increased difficulties with ACD resolution.

According to the second approach, the in-situ *wh*-phrase can be interpreted in-situ. Hence, encountering the in-situ *wh*-phrase will not trigger reanalysis that could facilitate downstream ACD resolution. Only upon reaching the ellipsis site itself will the need for reanalysis be apparent. Thus processing costs for ACD resolution are predicted to reflect both covert movement of the *wh*-object to a position above the missing VP as well as the retrieval of the appropriate antecedent for the elided VP and so should be relatively higher for both the *small* and *large* ellipsis conditions. Under this approach we thus expect to find a main effect of the object type, such that sentences with a relative clause headed by *which* are more difficult to process than sentences with a relative clause headed by *every*. This main effect may be accompanied by an interaction, such that the two *which* conditions pattern with *every-was* and are more difficult to process than *every-did*, or they may be even more difficult than the *every-was* condition.

(20) **Predictions with regard to ACD resolution in Experiment 1**

a. The covert movement approach:
The resolution of both *small* and *large* ellipsis is facilitated.
*Which-did* and *which-was* pattern with *every-did* and are easier to process than *every-was*.

b. The in-situ approach:
The resolution of both *small* and *large* ellipsis is not facilitated.
*Which-did* and *which-was* are more difficult to process than *every-did*; they are at least as difficult to process as *every-was*.

### 3.2.4 Results

61 native speakers of English participated in this study. The following exclusion criteria were used to filter the results of this experiment and all subsequent ones: participants who held the spacebar continuously pressed instead of reading the sentences one word at a time as instructed, participants who participated in the study more than once, participants who submitted the entire survey in less than 10 minutes,[10] participants with an average reaction time of over 700ms,[11] and participants with low accuracy rates in response to comprehension questions (<75% on filler trials and <75% on target trials) were excluded from the study. Twenty participants in Experiment 1 were excluded from the analysis for these reasons.[12] In addition, two target sentences and no filler sentences were excluded

[10] Average completion time for all of the experiments we present here was at least 25 minutes.

[11] Particularly slow reading times in our experiments tended to reflect long breaks and distracted behavior of the participants, introducing extraneous noise into the results. Particularly fast reading times were similarly contributed by distracted participants.

[12] Although this is exclusion rate is quite high, we believe that it is necessary in order to filter the relatively higher noise level in Amazon Mechanical Turk participants as compared to lab participants. We believe that

from the analysis because of low accuracy (<60% across participants).

Questions across the full experiment (targets and fillers) were answered correctly 87.5% of the time across participants; questions for experimental items were answered correctly on 83.3% of trials. The data was trimmed as follows: RTs from the first and last words of all items, RTs faster than 90ms or slower than 2000ms, and any RTs that were more than 2 standard deviations faster or slower than the average RTs for each subject (calculated per condition) were excluded from the analysis. Overall, less than 1% of the data were lost due to these criteria.[13] Figure 2 shows the mean residual reading times (RRTs)[14] for the two regions of interest for the four target conditions.

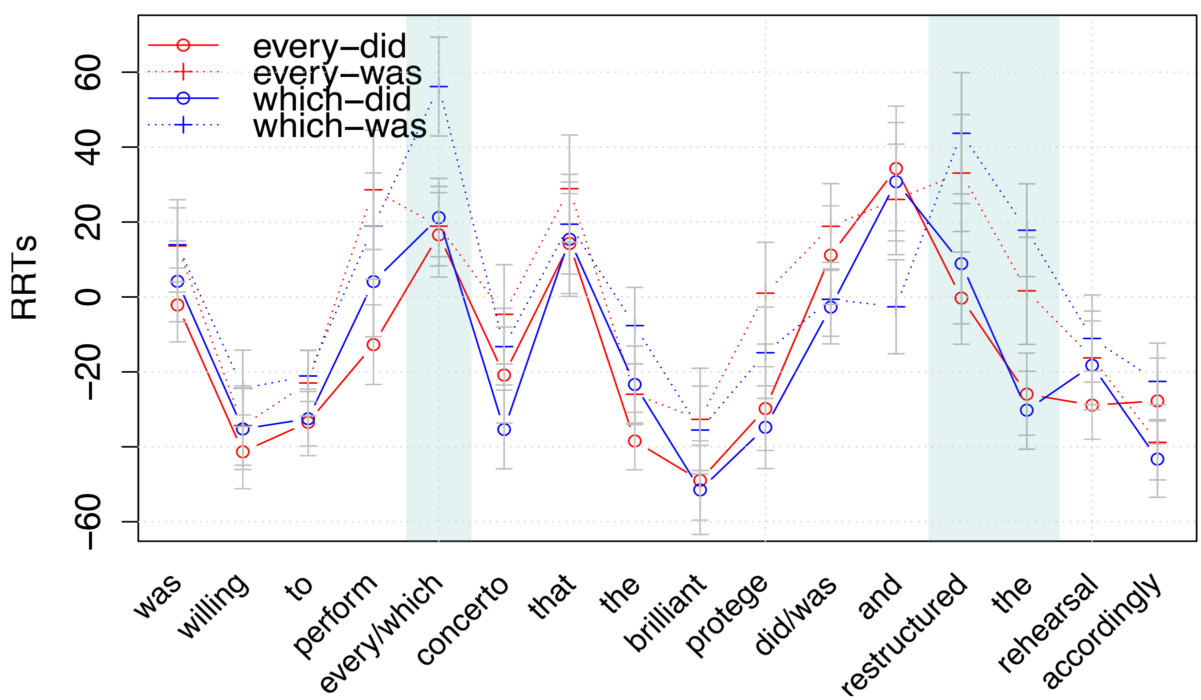



Figure 2: Residual reading times in target items in Experiment 1

A linear mixed effects model was fit to the data using *R* and the *R* package *lme4* (Bates and Sarkar, 2007). The model predicted RRTs from the two factors of interest: *determiner* (*every* vs. *which*) and *ellipsis size* (*small ellipsis* marked by did, vs. *large ellipsis* marked by was).[15] The model contained random intercepts and slopes for both predictors for subjects and items (Baayen, 2004; Barr et al., 2013).

The results show a main effect of *determiner* at the slot at which the determiner appeared in the sentence (log likelihood tests comparing a model with and without the effect of *determiner*, $p<0.05$). This result is driven by the fact that reaction times in the *which* condition

this more parsimonious criterion helps ensure the validity of the results we report here.

[13]Here and in the other experiments, the results remain statistically unaltered if this step is not performed.

[14]RRTs were calculated based on a regression equation predicting reading time from word length using all words from all experimental items, except for the first word and the last word of the sentence.

[15]Similar results are obtained for logRTs in our data. We report the results for RRTs for convenience.

were slower than the reaction times in the *every* condition, across both ellipsis conditions.[16] The results additionally show a main effect of *ellipsis size* two words and three words after the auxiliary site (log likelihood tests, $p$'s$<0.05$).[17] This result is driven by the fact that the resolution of *small ellipsis* is faster than the resolution of *large ellipsis* for both *every* and *which*. There were no differences between the two determiners at these slots, and there were no other significant effects in the results. The results of the model for the third word after the auxiliary site are summarized in Table 1.

| Predictor | Coefficient | Standard Error | $t$ value |
|---|---|---|---|
| Intercept | -31.591 | 18.851 | -1.676 |
| *Determiner* | 3.586 | 16.018 | 0.224 |
| *Ellipsis size* | 39.8416 | 15.651 | 2.546 |
| *Determiner*×*Ellipsis size* | 2.665 | 20.628 | 0.129 |

Table 1: Results of Experiment 1

### 3.2.5 Discussion

We find two effects in Experiment 1. First, the main effect of *determiner* in the first region of interest may be attributed to the relatively higher complexity of a multiple question compared to that of a simplex question. When the parser encounters the in-situ *wh*-phrase in the *which* conditions, it realizes that the sentence will be paired with a more complex semantics than in the case of the simplex question with *every*. Regardless of the cause of this effect, it shows that our participants were processing the sentences at least at a depth sufficient for detecting the difference in determiner.

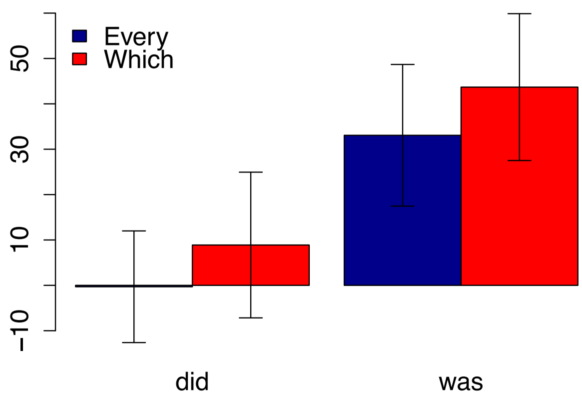



Figure 3: RRTs for 2nd word after AUX

Second, we find a main effect of *ellipsis size* in the second region of interest, following the auxiliary verb, such that sentences with a *small ellipsis* are read faster than sentences with a *large ellipsis*. This is the case for both determiners: *every* and *which*. To see this more clearly, observe Figure 3, which compares reading times for *every* and *which* two words after the auxiliary verb. As we can see, the *ellipsis size* manipulation affects the two determiners equally.

[16]Note, in particular, that the *which-was* condition appears to be slower than the other three conditions. However, the *determiner*×*ellipsis size* interaction is not significant ($p$=0.344).

[17]As we will see below, results consistently show up on the second and third words after the verb. This was furthermore confirmed for this experiment in a pilot experiment using the same items presented here. See also discussion in section 2.4.4.

This result is not predicted by either traditional approach to *wh*-in-situ. Recall that the covert movement approach predicts facilitation of both *small* and *large* ellipsis, because the in-situ *wh*-phrase must move to C for interpretation, preemptively undoing antecedent containment in both ellipsis conditions. We would thus expect that RTs for both *which* conditions, irrespective of whether they involve local or non-local ACD, would be lower than those for the *every* conditions. The in-situ approach, on the other hand, assumes that the *wh*-phrase can be interpreted without any movement at all, and hence predicts no facilitation by an upstream *which* for either the *small* or *large* ellipsis conditions. However, the results of Experiment 1 show an effect of ellipsis size but no effect of *determiner*, which is not explained under both approaches to *wh*-in-situ.

Based on the results of Hackl et al. (2012), we may take the fact that there is no difference between *every-did* and *which-did* to indicate that local ACD resolution is facilitated in both cases. More specifically, if *which* where a determiner that could be interpreted in situ like the we should have seen relatively longer RTs for *which-did*, just like Hackl et al. (2012) did for *every-did*. The fact that we didn't shows that *wh*-phrases behave like traditional quantificational determiners such as *every* with regard to ACD resolution in real time. This is compatible with a view under which *wh*-phrases behave as quantifiers that must covertly move (via QR) to the nearest propositional node for interpretation.[18] Similar proposals have been previously made by several researchers for a variety of languages (cf. Baker 1970; Dornisch 2000 for Polish, Huang 1995; Kim 1991 for Korean, Rullmann and Beck 1998). Once the *wh*-phrase has been integrated into the structure, it does not require any further movement: it may be interpreted in-situ at the first position in which it can be interpreted in the structure. We propose that the movement step assumed by the parser is the smallest that can produce an interpretable structure. We return to a discussion of this proposal below, but first we present Experiment 2, whose goal is to entertain an alternative interpretation of the results of Experiment 1.

### 3.3 Experiment 2a-b: *the* vs. *every* and *the* vs. *which*

Long distance dependencies, such as those required for the resolution of ellipsis, are constructed by integrating temporally and structurally distant linguistic material. As such, they require the support of working memory resources for their completion (Foraker and McElree, 2011). In order to establish a dependency, the parser must retrieve previously processed material from working memory at the gap position, and find an appropriate antecedent for the ellipsis. The distance between the gap site and the antecedent affects the retrieval process, resulting in longer reading times and decreased acceptability for longer dependencies (cf. Gibson, 1998; van Dyke and Lewis, 2003; Lewis et al., 2006; Warren and Gibson, 2002). Relatedly, we might expect semantically more complex antecedents (e.g. *willing to perform*) to be harder to retrieve from memory than simpler antecedents (e.g. *perform*).

[18]See Heim and Kratzer (1998) for a discussion of the type-mismatch problem with quantifiers in object position.

One possible interpretation of Experiment 1, then, is that the Hackl et al. (2012) paradigm was not successfully replicated. Instead, participants were not reading the sentences carefully and were not affected by the *determiner* manipulation at all. A related concern is that the experimental paradigm is not sensitive enough to detect differences between the processing of different quantifiers, perhaps because an embedding inside a question presents too much of a challenge to the participants. Hence, Experiment 1 only exhibits an effect of complexity of the antecedent, where the processing of long-distance ellipsis with a more complex antecedent is more costly than that of short-distance ellipsis with a simpler antecedent, and nothing more. To address this concern, first note that some comprehension questions in Experiment 1 directly targeted the relative clause, to ensure that it was processed at a deep level. Those questions did not suffer from lower accuracy rates than other comprehension questions. We furthermore detect a main effect of *determiner* in Experiment 1 once *which* and *every* are read, ensuring that participants did indeed pay attention to the determiners in the sentences.

However, to more directly address the concern that the Hackl et al. (2012) paradigm might not have worked as expected, Experiment 2a-b will attempt to replicate the Hackl et al. (2012) results—comparing the behavior of *every* and *the*—in the context of an embedded question, and then additionally compare the behavior of *which* and *the* in the same context. As we will see below, we find that ACD resolution with the quantificational determiner *every* is facilitated compared to ACD resolution in sentences with the non-quantificational definite article *the*, as predicted by the results of Hackl et al. (2012). We furthermore find that ACD resolution in sentences with *which* is facilitated compared to ACD resolution in sentences with the definite article *the*. This result will eliminate the concern that the results of Experiment 1 do not show any facilitation effects with *which*.

### 3.3.1 Design

To allow for a direct comparison with Experiment 1, Experiment 2 uses the same materials as in Experiment 1, with minor changes to accommodate the experimental manipulation of this experiment. Like Experiment 1, Experiment 2 presented participants with (embedded) *wh*-questions headed by a subject *wh*-phrase. Two factors were crossed: (a) *determiner*: whether the embedded question contained the quantificational determiner *every*, the *wh*-word *which*, or the definite article; and (b) *ellipsis size*: whether the sentence contained a *small ellipsis* marked by <u>*did*</u>, where the antecedent of the ACD site is the embedded $VP_2$, or *large ellipsis* marked by <u>*was*</u>, where the antecedent of the ACD site is the matrix $VP_1$. *Determiner* was treated as a between-subject factor: Experiment 2a compares *every* and *the*, and Experiment 2b compares *which* and *the*. The missing comparison, *which* and *every*, has already been done in Experiment 1. Sample items are given in (21)-(22):

(21) **Sample target item in Experiment 2a**

The conductor asked *which soloist* was [$_{VP1}$ willing to [$_{VP2}$ perform...

a. **every** concerto that the brilliant protégé <u>*did*</u> ... *(small ellipsis)*

b. **the** concerto that the brilliant protégé <u>*did*</u> ... *(small ellipsis)*
c. **every** concerto that the brilliant protégé <u>*was*</u> ... *(large ellipsis)*
d. **the** concerto that the brilliant protégé <u>*was*</u> ... *(large ellipsis)*
...and restructured the rehearsal accordingly.

(22) **Sample target item in Experiment 2b**
The conductor asked *which soloist* was [$_{VP1}$ willing to [$_{VP2}$ perform...
a. **which** concerto that the brilliant protégé <u>*did*</u> ... *(small ellipsis)*
b. **the** concerto that the brilliant protégé <u>*did*</u> ... *(small ellipsis)*
c. **which** concerto that the brilliant protégé <u>*was*</u> ... *(large ellipsis)*
d. **the** concerto that the brilliant protégé <u>*was*</u> ... *(large ellipsis)*
...and restructured the rehearsal accordingly.

The same 28 target sentences from Experiment 1 were used, along with the same 48 filler items. The only change introduced to the sentences was to the determiners, as indicated above. The comprehension questions to some items were also minimally changed to accommodate this manipulation. Experiment 2 used the same methods as described above for Experiment 1. See the Appendix for a full list of the materials.

#### 3.3.2 Predictions

We expect Experiment 2a to replicate the results of Hackl et al. (2012). In particular, we expect to see an effect of *ellipsis size*, such that *small* ellipsis is easier to process than *large* ellipsis for both *the* and *every*. In addition, we expect to find a difference in the processing of ACD in sentences with a relative clause headed by *every* and sentences with a relative clause headed by *the*, such that the processing of ACD in sentences with *every* is facilitated compared to sentences with the.

In the *every* conditions (21a,c), the parser must assume covert movement as soon as it encounters *every*. This movement is expected to be local and to target a position above the embedded $VP_2$ headed by *perform* in (21), a position that is high enough to preemptively undo antecedent containment in the case of *small* ellipsis (21a) but not in the case of the *large* ellipsis (21c). Hence, the processing costs at the ellipsis site in the *small ellipsis* condition (22a) should be lessened since part of the work to resolve ACD, specifically reanalysis to undo antecedent containment, has already happened prior to encountering the ACD site. When the parser encounters <u>*did*</u> all it needs to do is find an antecedent for the missing VP and nothing more. In the case of the *large* ellipsis in (21c), the covert movement that was assumed following the processing of *every* is not sufficient: after local movement of the object DP the missing VP is still contained inside its antecedent. Hence, when the auxiliary <u>*was*</u> is reached, the parser must again reanalyze the structure, covertly moving the object a second time, from its position above the embedded $VP_2$ to a position higher than the matrix $VP_1$. Hence, a high processing cost should be associated with the resolution of the *large ellipsis* in (21c).

In the *definite* conditions in (21b,d), no covert movement is predicted to take place when the definite article is encountered. The parser only assumes a structure with covert movement of the object when it encounters the auxiliaries *did* and *was*, as the ACD marker is the earliest point at which the need for moving the object DP is detectable. The fact that processing the ACD site requires both reanalysis and retrieval of an antecedent is expected to result in relatively high processing costs following the ellipsis site. Moreover, just like in the case of *every*, we expect that the processing costs for the *large ellipsis* condition should be higher than those associated with the *small ellipsis*, since retrieving a more complex antecedent VP is more difficult and requires longer distance covert movement than in the case of the small ellipsis.

We furthermore expect to find the same behavior pattern with *which* in Experiment 2b, (22): if *which*, like *every*, is a quantifier in object position that must QR to the first propositional node in the structure for interpretation, we expect to find that *which*, like *every*, facilitates the resolution of small ellipsis with *did*, but not of large ellipsis with *was*, for the same reasons as described above for *every*. We expect the not to facilitate ACD resolution of any size, and hence we predict that sentences with the to be relatively more difficult to process than sentences with *which*.

#### 3.3.3 Results

165 native speakers of English participated in this study: 84 subjects participated in Experiment 2a and 81 participated in Experiment 2b. 21 subjects were excluded from the analysis of Experiment 2a and 24 subjects were excluded from the analysis of Experiment 2b using the same exclusion criteria specified in Experiment 1. Three target sentences and one filler sentence were excluded from the analysis of Experiment 2a and three target sentences and two filler sentence were excluded from the analysis of Experiment 2b because of low accuracy (<60% across all participants).

Questions across the full experiment (targets and fillers) were answered correctly 86.7% of the time across participants in Experiment 2a and 85.8% of the time in Experiment 2b; questions for target items were answered correctly on 84.7% of trials in Experiment 2a and 83.8% of the time in Experiment 2b. The data was trimmed using the same criteria described for Experiment 1. Overall, less than 1% of the data was excluded from the analysis. Figure 4 shows the mean residual reading times for the region of interest for the four target conditions in Experiment 2a, comparing the processing of sentences with *the* and *every*.

A linear mixed effects model with random intercepts and slopes for *ellipsis size* for subjects and items was fit to the data.[19] The model predicted RRTs from the two factors of interest: *determiner* (*every* vs. *the*) and *ellipsis size* (*small ellipsis* marked by *did*, vs. *large ellipsis* marked by *was*). The results show a main effect of *ellipsis size* two and three words after the auxiliary verb site and a main effect of *determiner* two words after the auxiliary verb

[19] A more specified model that includes the effect of *determiner* yields a false convergence.

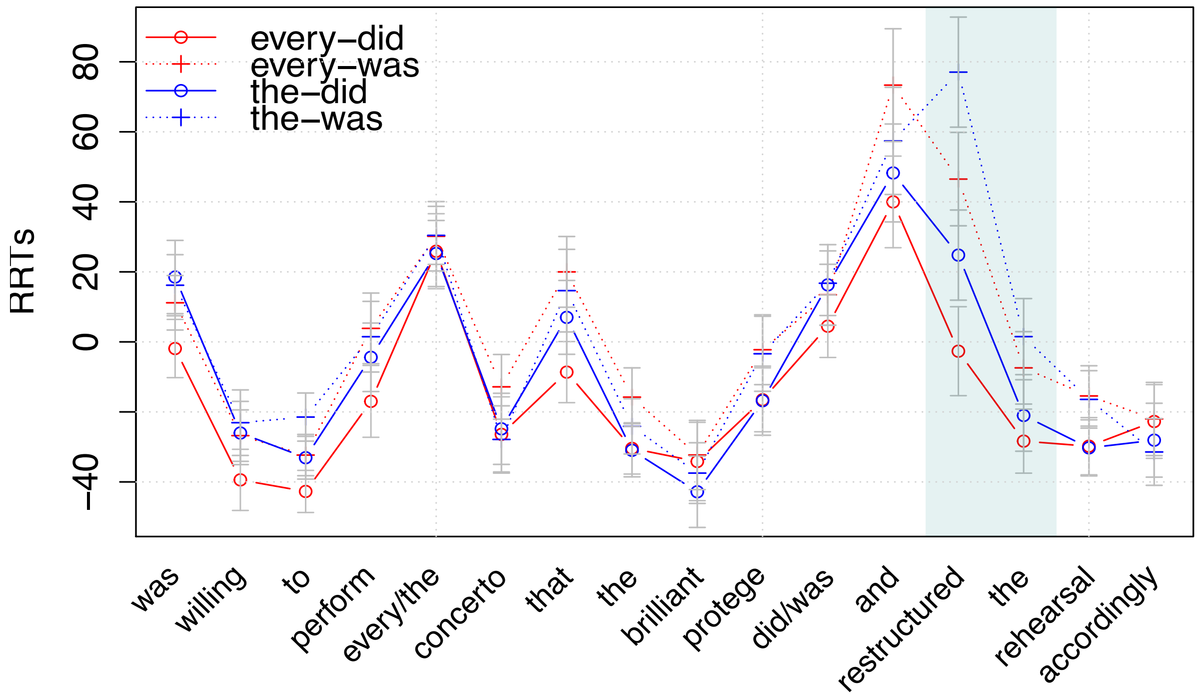


Figure 4: Residual reading times in target items in Experiment 2a

site (log likelihood tests comparing a model with and without the predictor of interest, $p$'s$<0.05$). These results are driven by the fact that the resolution of *small ellipsis* is faster than the resolution of *large ellipsis* for both every and the, and furthermore that the resolution of ACD in sentences with a relative clause headed by *every* is faster than the resolution of ACD in sentences with a relative clause headed by *the*. There were no other significant effects in the data. The results of the model for the second word after the auxiliary site are summarized in Table 2.

| Predictor | Coefficient | Standard Error | $t$ value |
|---|---|---|---|
| Intercept | 3.304 | 20.959 | 0.158 |
| *Determiner* | 21.780 | 16.517 | 1.319 |
| *Ellipsis size* | 39.008 | 19.510 | 1.999 |
| *Determiner*×*Ellipsis size* | 12.207 | 23.441 | 0.521 |

Table 2: Results of Experiment 2a

Next, we examine the results of Experiment 2b, comparing the determiners *the* and *which*. Figure 5 shows the mean RRTs for the region of interest for the four target conditions.

A linear mixed effects model with random intercepts and slopes for *ellipsis size* for subjects and items was fit to the data. The model predicted RRTs from the two factors of interest: textitdeterminer (*which* vs. *the*) and *ellipsis size* (*small ellipsis* marked by <u>did</u>, vs. *large ellipsis* marked by <u>*was*</u>). The results show a main effect of *ellipsis size* two and three words after the auxiliary verb site and a main effect of *determiner* two words after the auxiliary verb

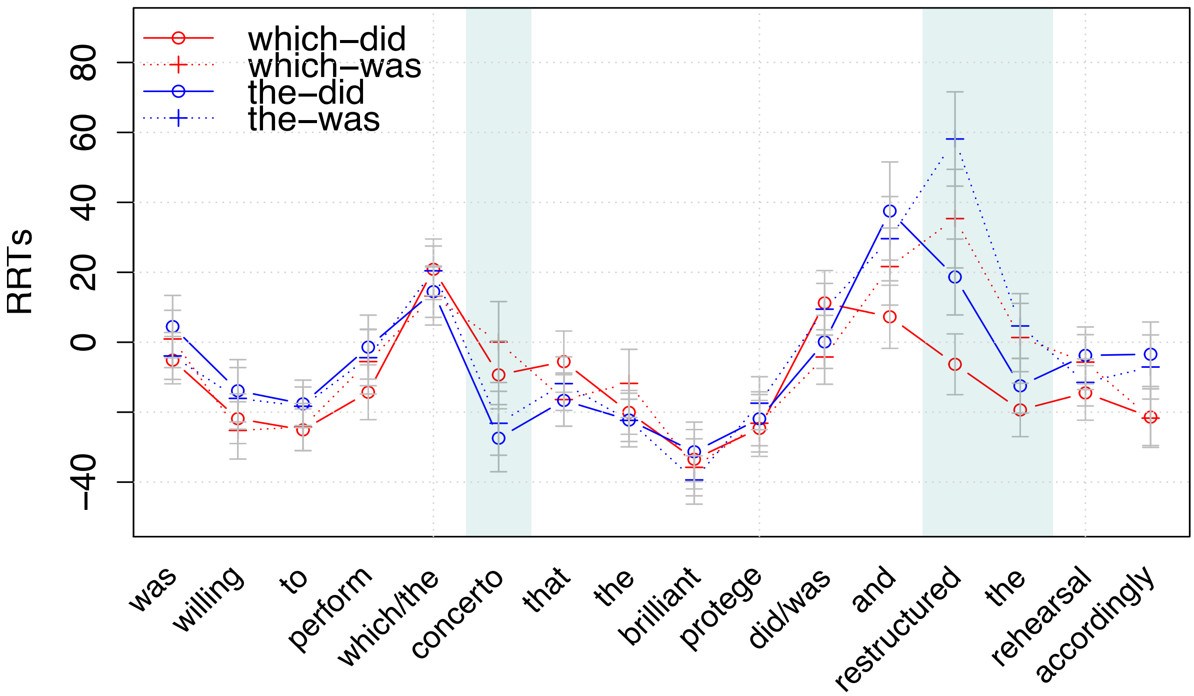


Figure 5: Residual reading times in target items in Experiment 2b

site (log likelihood tests comparing a model with and without the predictor of interest, $p$'s$<0.05$). These results are driven by the fact that the resolution of *small ellipsis* is faster than the resolution of *large ellipsis* for both *which* and *the*, and furthermore that the resolution of ACD in sentences with a relative clause headed by *which* is faster than the resolution of ACD in sentences with a relative clause headed by *the*. There were no other significant effects in the data. The results of the model for the second word after the auxiliary site are summarized in Table 3.

| Predictor | Coefficient | Standard Error | $t$ value |
|---|---|---|---|
| Intercept | 16.572 | 15.716 | 1.054 |
| *Determiner* | -22.748 | 14.213 | -1.600 |
| *Ellipsis size* | 42.714 | 15.949 | 2.678 |
| *Determiner*×*Ellipsis size* | -5.117 | 20.483 | -0.250 |

Table 3: Results of Experiment 2b

### 3.3.4 Discussion

The results of Experiment 2a-b confirm that the Hackl et al. (2012) paradigm extends to the context of an embedded question. In particular, we find—in addition to the main effect of *ellipsis size* also a main effect of *determiner*, such that sentences with *every* are processed faster than sentences with *the*, and sentences with *which* are also processed faster than sentences with *the*. That is, *every* and *which* pattern together and facilitate ACD resolution more than parallel sentences with the. To see this more clearly, consider Figure 6, which compares the reading times of *the*, *every* and *which* two words after the auxiliary verb site in Experiments 2a-b.

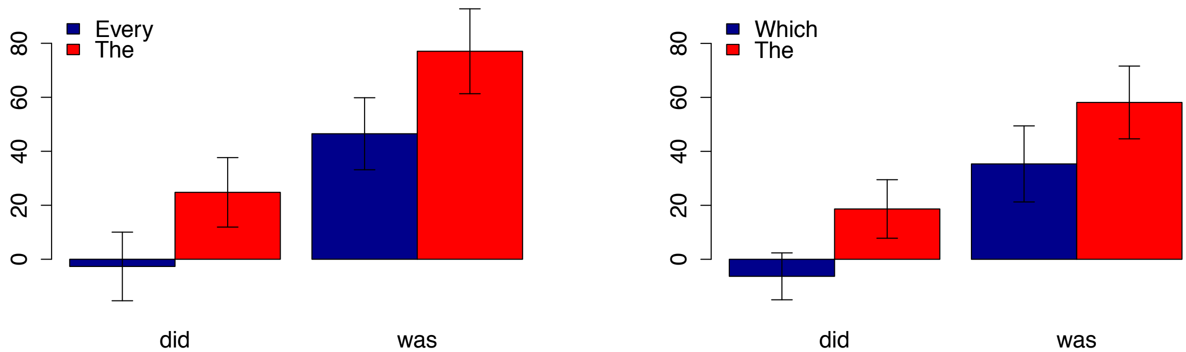



Figure 6: RRTs two words after AUX in Experiments 2a (left) and 2b (right)

For the results of Experiment 2a, we adopt Hackl et al.'s (2012) explanation and assume that this is so because the parser must assume covert movement with *every* as soon as the quantifier is encountered, but no such covert movement is assumed with *the*.[20] As a result, covert movement must always be assumed to take place at the point of ACD resolution in the case of *the* but not in the case of *every*: when a *small ellipsis* (marked with *did*) is encountered, the parser must assume no additional covert movement with *every* since that has happened earlier in the parse. To explain the fact that non-local ACD resolution is less difficult with *every* as well we hypothesize that the second step of QR to a position above the matrix $VP_1$ is easier in the case of *every*, perhaps because the object needs to be moved less far than in the case of *the*.

The results of Experiment 2b are similarly explained if we assume that *wh*-phrases behave as quantifiers that must QR to the nearest propositional node for interpretation as soon as they are encountered by the parser, but that no such movement is assumed in the case of the definite article *the*. As a result, covert movement must always be assumed to take place at the point of ACD resolution in the case of *the* but not in the case of *which*: when a *small ellipsis* (marked with *did*) is encountered, the parser must assume no additional covert movement with *which*, since sufficient movement has already happened earlier in the parse. As in the case of *every* in Experiment 2a, we find that non-local ACD resolution is less difficult with *which* than with *the*. This is explained in the same way as for *every*: we hypothesize that the second step of QR to a position above the matrix $VP_1$ is easier in the case of *which* because the object needs to be moved less far than in the case of *the*.

[20]However, an important difference between our results and Hackl et al.'s (2012) results is that while Experiment 2 yields two main effects, Hackl et al. observed an interaction, such that *every-was* and *the-was* were not significantly different from each other. This played an role in Hackl et al.'s argument that only local QR facilitates ACD resolution. Below, the results of Experiment 3 will provide us with the missing evidence to make this claim in this paper here as well. We return to this point after introducing that experiment.

## 3.4 General discussion

The results of Experiment 2 thus confirm that Experiment 1 is sensitive enough to detect potential differences between determiners with regard to their interaction with ACD, and that there is no reason to suspect that participants were not processing the experimental materials with sufficient depth. Here again, we find similar accuracy rates on target and filler sentences as in Experiment 1, and the overall reaction times in Experiment 2 are comparable to those found in Experiment 1 (see Figure 2).

The facilitation effect found with *every* compared to the definite article *the* and the fact that parallel behavior is observed with *which* support the interpretation we have given to the results of Experiment 1. Specifically, they show that local ACD resolution with *which* is facilitated just as it is for *every*, and both determiners are easier to process than *the*.

(23) **The results of experiments 1-2**

a. *Small* ellipsis (*did*) is easier to process than *large* ellipsis (*was*), regardless of the determiner (*which*, *every*, *the*) heading the relative clause hosting the ellipsis.

b. When the relative clause hosting the ellipsis is introduced by either *which* or *every*, it is easier to process than if it is introduced by *the*.

The facilitation effect with *which* is not compatible with the *wh*-phrase having been processed in-situ when it was encountered by the parser, as we would then predict no effect of *determiner* in Experiment 2b.

(24) **Against the in-situ approach to *wh*-in-situ**
The results of Experiments 1-2 are incompatible with an LF in which the in-situ *wh* in an English multiple question undergoes no covert movement.
Surface in-situ *wh*-phrases undergo covert *wh*-movement.

The results are compatible with a covert movement approach to *wh*-in-situ. However, recall that such an approach traditionally assumes that covert movement must target interrogative C. The results of Experiments 1 and 2b are incompatible with this state of affairs: we observe that *which* patterns with *every*, and importantly the *which-was* condition is not easier to process than the *every-was* condition. If covert *wh*-movement is long-distance, whereas movement with *every* is local, we would expect to find a difference between these two conditions. Instead, the results are consistent with *which* being QRed *locally* to a position above the embedded $\mathrm{VP}_2$, leading to a prediction that *which* should facilitate ACD resolution to a similar extent as *every*. This is indeed the result of Experiments 1 and 2b. This leads us to propose a new approach to the interpretation of *wh*-in-situ, the *partial movement* approach. Below we present Experiment 3, which provides further support to this view of *wh*-in-situ.

(25) **The *partial movement* approach**
A *wh*-phrase must be interpreted at a position with propositional type at LF.

# 4 *Intervention effects* and the nature of covert movement

We concluded from Experiments 1-2 that (a) surface in-situ *which* is not interpreted in-situ, and moreover (b) surface in-situ *which* covertly moves only a short distance, similar to *every*. Experiment 3 will provide support for the latter conclusion by considering a prediction of the partial movement approach: that although *wh*-phrases need not move any further once they have been integrated into the structure, they may be forced to move higher on independent grounds. This prediction allows us to distinguish between the behavior of *every* and the behavior of *which* in certain environments, where long-distance *wh*-movement may be necessary, leading us to expect a larger region in the question in which ACD facilitation effects are expected with *which* but not with *every*. Below we test this prediction using *intervention effects* (Beck, 2006, among others) and show that it is indeed borne out.

(26) **Predictions of the partial movement approach**
*Wh*-phrases QR to the nearest propositional node for interpretation.
No additional movement is required for the interpretation of a *wh*-phrase,
*but movement may be forced by external interpretability considerations*.

## 4.1 The theory of intervention effects

The *partial movement* approach to *wh*-in-situ predicts that, although *wh*-phrases in object position can be interpreted in-situ following an initial QR step, other factors may force further *wh*-movement. If long-distance movement can be forced, we expect to find facilitation effects of ACD resolution in a larger portion of the sentence, up to the position targeted by covert movement. One phenomenon that has been argued to force covert movement of in-situ *wh*-phrases is *intervention effects*. Importantly, this phenomenon has not been argued to affect quantificational determiners like *every*, and so intervention effects allow us to make diverging predictions about the behavior of *every* and *which* and to experimentally test these predictions.

The term *intervention effect* describes a situation in which a question is rendered ungrammatical because an in-situ *wh*-phrase is c-commanded at LF by an *intervener*, for instance a focus sensitive operator such as *only* or negation. Cross-linguistically, intervention effects have been found in *wh*-fronting languages as well as in *wh*-in-situ languages, and several competing theories have been proposed to explain the phenomenon (cf. Beck, 1996, 2006; Beck and Kim, 1997; Kim, 2002; Pesetsky, 2000; Tomioka, 2007a,b; Cable, 2010; Mayr, to appear, see also Hagstrom 1998; Hoji 1985; Soh 2005).

Here we adopt Beck's (2006) approach to intervention effects, who, following Pesetsky (2000), proposes that these effects are caused whenever a *wh*-phrase cannot covertly move above a potentially offending intervener and is instead forced to be interpreted in-situ using focus- alternatives. The intervener is argued to disrupt the transmission of alternatives from the *wh* to C, resulting in ungrammaticality, (27b). If, on the other hand, the *wh* is able to covertly move to a position above the intervener, the question can be assigned a convergent semantics, (27a). This state of affairs is summarized in the schema in (27), where covert movement is marked with an arrow and areas where alternatives are projected are marked with a curly arrow.[21]

(27) **The configuration of an intervention effect (Beck, 2006)**

a. ✓ [$_{CP}$ C ... *wh* ... **intervener** ... *t* ... ]

b. * [$_{CP}$ C ... **intervener** ... *wh* ... ]

In English, superiority-obeying questions are immune to intervention effects, but superiority-violating questions are ungrammatical whenever an intervener occurs above an in-situ *wh*-phrase (É Kiss, 1986; Pesetsky, 2000). This is illustrated in (28a-b), which employ the focus-sensitive operator ***only*** as an intervener.

(28) **Intervention effects only target superiority-violating questions**

a. *Which student* did **only** Fred introduce __ to *which professor*? Sup.-obeying

b. * *Which professor* did **only** Fred introduce *which student* to __? Sup.-violating

Pesetsky (2000) and Beck (2006) argue that in superiority-obeying questions, the in-situ *wh* is able to covertly move to C at LF. As a result, it is above the offending intervener at LF and hence the question is grammatical, (28a). In superiority-violating questions, on the other hand, the in-situ *wh* cannot move to C and must stay in-situ at LF and be interpreted via focus-alternatives. As a result, if an intervener occurs above the in-situ *wh*-phrase, the result is ungrammaticality, (28b).

The grammaticality of example (28a) is explained because the in-situ *wh*-phrase *which professor* is able to covertly move above the intervener at LF. This movement of *which professor* targets a higher position than what is expected if no intervener is present. This is shown in (29)-(30) below, where a question with and without an intervener are contrasted. We can see that longer movement has occurred in the question with an intervener (30) than in the question that lacks an intervener, (29): In example (29), *which professor* has covertly moved above the predicate *introduce* to satisfy the interpretability requirement which forces it to QR to a propositional node for interpretation. In example (30), *which professor* undergoes covert movement above the intervener (and perhaps as far as C) in order to avoid the illicit intervention effect configuration in (27b). If *which professor* were not covertly raised above the intervener, the result would be the ungrammatical (30a) (which is identical to (29) except for the presence of the intervener). As before, dashed arrows represent covert movement and curly arrows represent areas in which focus-alternatives are computed. For simplicity, we do not draw arrows to represent the overt movement in these LFs.

[21] For Beck (2006), covert movement necessarily targets the interrogative complementizer. That is, Beck adopts the traditional all-or-nothing view of covert movement. However, in principle all that is necessary for the *wh*-phrase to be interpretable in Beck's theory is for it to move above the intervener. This is compatible with a smaller movement step under the partial movement approach, as schematized in (27a).

(29) **LF of multiple question under the partial movement approach**
*Which student* did Fred introduce ___ to *which professor*?
[[*which student*]$_1$ [ C [$_{TP}$ ... t$_1$ ... [*which professor*]$_2$ [$_{VP}$ introduce ... t$_2$ ]]]]

(30) **LF of question with intervener under the partial movement approach**
*Which student* did **only** Fred introduce ___ to *which professor*?

a. * [[*which student*]$_1$ [C [$_{TP}$ ... **intervener** t$_1$ [*which professor*]$_2$ [$_{VP}$ introduce ... t$_2$]]]]

b. [[*which student*]$_1$ [C [$_{TP}$ ... [*which prof.*]$_2$ **intervener** t$_1$ t$_2$ [$_{VP}$ introduce ... t$_2$]]]]

We thus predict that the presence of an intervener in a question can force additional covert movement of *wh*-phrases that could otherwise be interpreted without movement following their QR step. Hence, we expect the presence and placement of an intervener in a question to affect the ease of ACD resolution in that question. The intervener marks the scope of the necessary covert movement in the question: the *wh*-phrase must move at least as high as the intervener in order to escape intervention. We thus expect to find facilitation of ACD resolution if the ellipsis is smaller than the movement required by the intervener. On the other hand, no additional movement step is predicted to occur in the case of the quantificational determiner *every*, because interveners do not interact with traditional quantifiers. Consequently, no additional facilitation of ACD resolution is expected when an intervener occurs in the context of *every*. Testing this prediction will be the goal of Experiments 3a-b.

Experiment 3a-b will use the focus-sensitive intervener ***also***. Although previous work on English has not used this intervener, it is known to be focus-sensitive (Beaver and Clark, 2008, a.o.) and has been argued to act as an intervener in other languages (Kim, 2002; Beck, 2006, a.o.). We argue that it acts as an intervener in English as well, exhibiting the same grammaticality paradigm as known interveners such as ***only*** and ***even***, where a superiority-obeying question with ***also*** above the in-situ *wh*-phrase is grammatical, (32a), but a superiority-violating question with ***also*** above the in-situ *wh*-phrase is ungrammatical, (32b, cf. example (28)). Compare this with the baselines in (31), which are minimally different in that they do not contain an intervener. Here, both the superiority-obeying question (31a) and the superiority-violating question (31b) are grammatical.[22]

(31) **Baseline: sup.-obeying and sup-violating questions both grammatical**
I know that the teacher punished some students last week. Do you know...

a. *which punishment* the teacher gave ___ to *which student* on Friday?

b. *which student* the teacher gave *which punishment* to ___ on Friday?

[22]Judgments regarding intervention effects are notoriously delicate. The question must be read with a supporting context, so that it receives a pair-list interpretation and not a single-pair interpretation. Many speakers additionally find the judgment crisper in an embedded context compared to a matrix question.

(32) ***Also* is an intervener in English**

I know that the teacher punished some students on Thursday. Do you know...

a. *which punishment* the teacher ***also*** gave __ to *which student* on Friday?

b. * *which student* the teacher ***also*** gave *which punishment* to __ on Friday?

## 4.2 Experiments 3a-b: *every* vs. *which* with interveners

### 4.2.1 Design

Experiment 3 is an expansion of Experiment 1, adding interveners to the questions. Three factors will be crossed: (a) *determiner*: whether the embedded question contained the quantificational determiner *every*, yielding a simplex *wh*-question, or a second *wh*-phrase, yielding a multiple *wh*-question; (b) *ellipsis size*: whether the sentence contained a *small ellipsis* marked by <u>did</u>, where the antecedent of the ACD site is the embedded $VP_2$, or *large ellipsis]* marked by <u>was</u>, where the antecedent of the ACD site is the matrix $VP_1$; and (c) *position of also*: whether the intervener *also* is *low* and occurs above the embedded $VP_2$, or *high* and occurs above the matrix $VP_1$.

To simplify the design and the analysis of the results, Experiment 3 is divided into two sub-experiments, each using just one determiner, *every* or *which*. Sample items for Experiments 3a-b are given in (33)-(34). See the Appendix for a full list of the materials.

(33) **Sample target item in Experiment 3a (*every*)**[23]

The conductor asked *which soloist* was ...

a. [$_{VP1}$ **also** willing to [$_{VP2}$ perform every concerto that the brilliant protégé <u>*did*</u> ... *(sm ellipsis)*

b. [$_{VP1}$ willing to [$_{VP2}$ **also** perform every concerto that the brilliant protégé <u>*did*</u> ... *(sm ellipsis)*

c. [$_{VP1}$ **also** willing to [$_{VP2}$ perform every concerto that the brilliant protégé <u>*was*</u> ... *(lg ellipsis)*

d. [$_{VP1}$ willing to [$_{VP2}$ **also** perform every concerto that the brilliant protégé <u>*was*</u> ... *(lg ellipsis)*

...and restructured the rehearsal accordingly.

(34) **Sample target item in Experiment 3b (*which*)**

The conductor asked *which soloist* was ...

a. [$_{VP1}$ **also** willing to [$_{VP2}$ perform which concerto that the brilliant protégé <u>*did*</u> ... *(sm ellipsis)*

b. [$_{VP1}$ willing to [$_{VP2}$ **also** perform which concerto that the brilliant protégé <u>*did*</u> ... *(sm ellipsis)*

c. [$_{VP1}$ **also** willing to [$_{VP2}$ perform which concerto that the brilliant protégé <u>*was*</u> ... *(lg ellipsis)*

d. [$_{VP1}$ willing to [$_{VP2}$ **also** perform which concerto that the brilliant protégé <u>*was*</u> ... *(lg ellipsis)*

...and restructured the rehearsal accordingly.

[23] We note that there are several choices for what element in (33)-(34) might be chosen by a reader to be the associate of *also*. One natural choice is *protégé*, which contrasts with the *soloist* occurring higher in the sentence. Other options native speakers have reported to us are possible choices are *concerto* and the verb *perform*. The actual choice made by readers is immaterial to our predictions for this experiment: all that matters is that *also* acts as a focus-sensitive operator, forcing movement above it in the case of *which* but not in the case of *every*.

Experiment 3 contained the same 28 target sentences from Experiment 1, along with the same 48 filler items. The items were minimally changed to add the intervener *also* in the appropriate places. No changes were made to the comprehension questions.

### 4.2.2 Predictions

As sketched above, we expect interveners to interact with *wh*-phrases and force movement of the in-situ *wh*-phrase to a position above the intervener. Following our assumptions about the parser, we expect it to perform the smallest movement possible, that is to target the first interpretable position above the intervener and no higher. Hence, in the *low also* condition, we expect *wh*-movement to target a position above *also* but below $VP_1$, while in the *high also* condition we expect long distance *wh*-movement above also and thus above $VP_1$. We do not expect interveners to interact with *every*, as conventional quantifiers do not interact with focus interveners, and hence we expect *every* to undergo a small QR step and not to move any further until the need for long-distance movement is apparent after reaching the ACD site marked by *was*.

For Experiment 3a we expect to replicate the results of Experiment 1, since interveners should not affect the parsing of sentences with *every*. That is, we expect to find a main effect of *ellipsis size* and, crucially, no interaction with the *placement of also*. The parser assumes covert movement as soon as it encounters *every* to a local position just above the embedded $VP_2$. This position is high enough to preemptively undo antecedent containment in the case of *small ellipsis* (33a-b) but not in the case of the *large ellipsis* (33c-d). Hence, no particular processing cost should be associated with the *reanalysis* step of ACD resolution at the ellipsis site in the *small ellipsis* conditions, because no additional movement is required at the gap site in order to resolve ACD in this sentence. All that the parser must do is identify an appropriate antecedent for the ellipsis, fill it into the gap and compute the resulting meaning. In the case of the *large ellipsis*, however, the covert movement that was assumed following the processing of *every* is not sufficient: following this movement, the missing VP is still contained inside its antecedent. Hence, when the auxiliary *was* is reached, the parser must again reanalyze the structure, covertly moving the object a second time from its position above the embedded $VP_2$ to a position higher than the matrix $VP_1$. Hence, a high processing cost should be associated with the resolution of the *large ellipsis* in (33c-d). Since the presence and position of *also* should not affect the movement of *every*, we predict no differences between the *low also* conditions (33b,d) and *high also* conditions (33a,c).

Different predictions are made in the case of Experiment 3b. Here, we expect the position of *also* to affect the amount of covert movement that the in-situ *wh*-phrase must undergo irrespective of the size of ACD.

In the *low also* conditions (34b,d), we expect the results to resemble those of Experiment 1 and of Experiment 3a. The position of *also* above the embedded $VP_2$ will force local movement, targeting a position above *also* in the embedded $VP_2$ but below the matrix $VP_1$. This position is high enough to preemptively undo antecedent containment in the case of *small ellipsis* (34b) but not in the case of the *large ellipsis* (34d). Consequently, we expect ACD resolution to be facilitated in the case of *small ellipsis* but not in the case of the *large ellipsis*, resulting in larger RRTs for the *was* conditions compared to the *did* condition.

In the *high also* conditions (34a,c), the position of *also* is above the matrix $VP_1$. As always, the parser must assume covert movement as soon as it encounters *which*. Unlike in the case of *low also*, here the movement must be non-local and target a position above the matrix $VP_1$. This position is high enough to preemptively undo antecedent containment both in the case of *small ellipsis* (34a) and *large ellipsis* (34c). As a result, when the parser reaches the gap site—in both ellipsis conditions—no reanalysis is necessary in order to construct an appropriate antecedent for the ellipsis. All the parser must do is find an antecedent for the ellipsis in the already constructed structure, fill it into the gap and compute the resulting meaning, but nothing more. We thus expect ACD resolution to be facilitated in both the *small ellipsis* and the *large ellipsis* conditions.

In summary, in addition to a main effect of *ellipsis size* which we expect to find in both Experiments 3a-b, we expect no effect of the *position of also* in the case of Experiment 3a (with *every*), but we do expect an effect of the *position of also* in Experiment 3b (with *which*). In particular, a *high also* should force a longer covert *wh*-movement step than a *low also*, and this additional movement is expected to create a larger region in which the facilitation of ACD resolution is expected. Both experiments contain conditions that will serve to replicate the results of the previous experiments. Furthermore this experiment will present novel data on the online processing of questions with interveners.

### 4.2.3 Results

243 native speakers of English participated in this study: 123 subjects participated in Experiment 3a and 120 subjects participated in Experiment 3b.[24] 27 subjects were excluded from the analysis of Experiment 3a and 25 subjects were excluded from the analysis of Experiment 3b using the same exclusion criteria specified in Experiment 1. Two target sentences and three filler sentences were excluded from the analysis of both experiments because of low accuracy (<60% across all participants).

Questions across the full experiment (targets and fillers) were answered correctly 84.9% of the time across participants in Experiment 3a and 86% of the time across participants in Experiment 3b; questions for experimental items were answered correctly on 82.42% of trials in Experiment 3a and on 82.47% of the trials in Experiment 3b. The data was trimmed as described for Experiment 1 above. Overall, less than 1% of the data was excluded. Below we present first the results of Experiment 3a and then those of Experiment 3b. Figure 7 shows the mean RRTs for the region of interest for the four target conditions in Experiment 3a. Recall that Experiment 3a contained *every* as the determiner heading the object relative clause in all target sentences.

[24] The details for Experiments 3a-b represent the aggregate results from two separate replications of each experiment, requested by an anonymous reviewer in order to ensure that the results are indeed reliable.

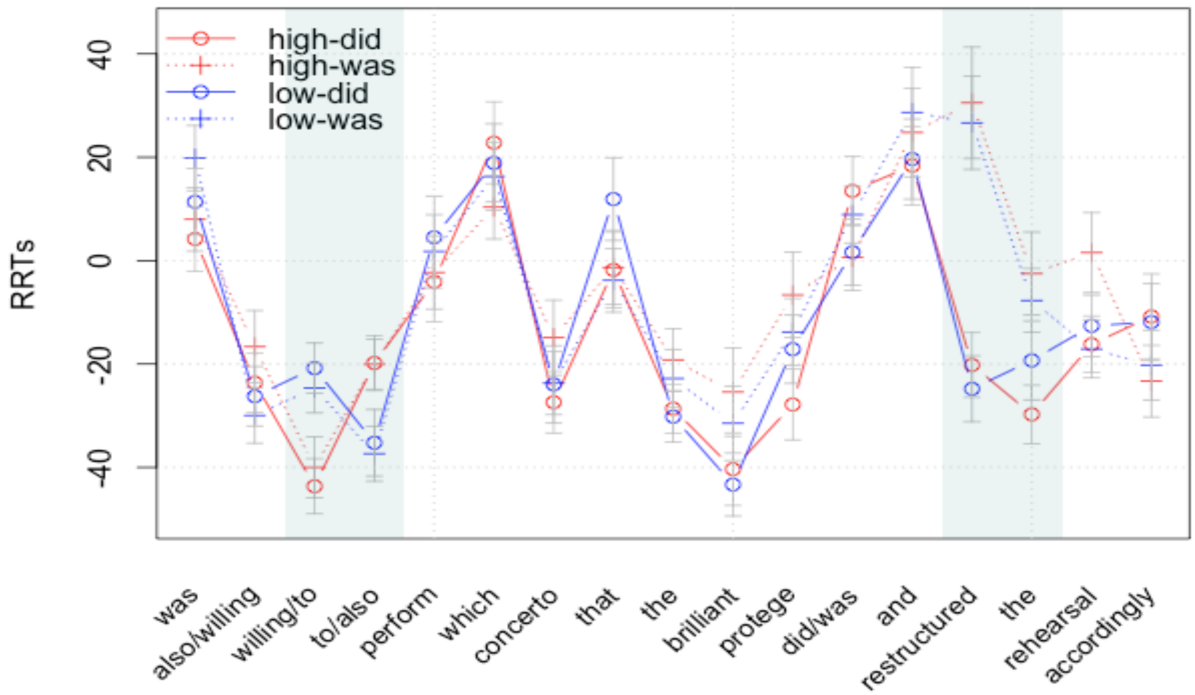



Figure 7: Residual reading times in target items in Experiment 3a

A linear mixed effects model with random intercepts and slopes for subjects and random intercepts and slopes for the effect of *ellipsis size* for items was fit to the data[25]. The model predicted RRTs from the two factors of interest: *position of also* (*low*, above the embedded verb vs. *high*, above the matrix verb) and *ellipsis size* (*small ellipsis*, marked by *did*, vs. *large ellipsis*, marked by *was*). The results show a main effect of *position of also* at the third and fourth word in the region of interest (where *also* occurs in the sentences). The results additionally show a main effect of ellipsis size two and three words after the auxiliary site (log likelihood tests comparing a model with and without the predictor of interest, $p$'s$<0.05$). These results are driven by the fact that the resolution of *small ellipsis* is faster than the resolution of *large ellipsis* for both the *low also* and *high also* conditions. There were no other significant effects in the data. The results of the model for the third word after the auxiliary site are summarized in Table 4.

| Predictor | Coefficient | Standard Error | $t$ value |
|---|---|---|---|
| Intercept | -28.342 | 11.909 | -2.380 |
| *Position of also* | 7.449 | 8.282 | 0.900 |
| *Ellipsis size* | 25.073 | 10.653 | 2.354 |
| *Position of also*×*Ellipsis size* | -14.070 | 13.034 | -1.079 |

Table 4: Results of Experiment 3a

[25] A more specified model with slopes for the effect of *also* for items did not converge.

Next, we examine the results of Experiment 3b. Recall that in this experiment, the determiner heading the object relative clause in all four target conditions was *which*. Figure 8 shows the mean RRTs for the region of interest for the four target conditions.

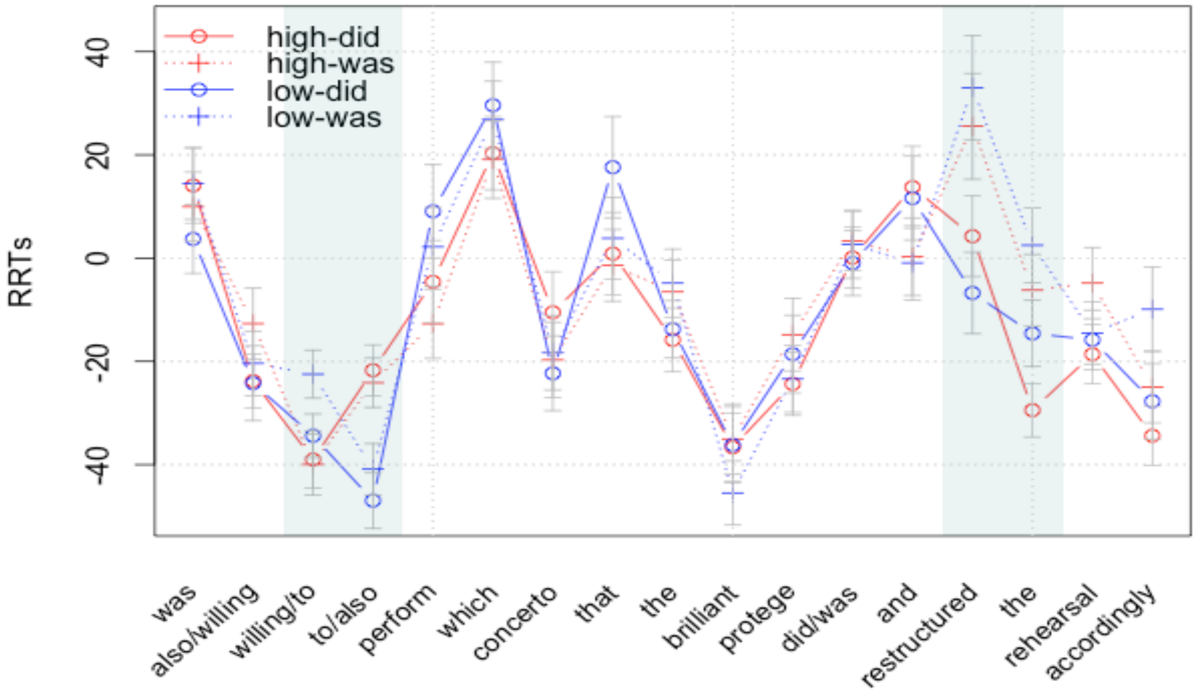



Figure 8: Residual reading times in target items in Experiment 3b

A linear mixed effects model with random slopes and intercepts for subjects and items was fit to the data. The model predicted RRTs from the two factors of interest: *position of also* (*low*, above the embedded verb vs. *high*, above the matrix predicate) and *ellipsis size* (*small ellipsis*, marked by *did*, vs. *large ellipsis*, marked by *was*). The results show a main effect of *position of also* at the third and fourth word in the region of interest (where *also* occurs in the sentences). The results additionally show a main effect of *ellipsis size* two and three words after the auxiliary verb site and a main effect of the *position of also* three words after the auxiliary site (log likelihood tests, $p$'s<0.05). The main effect of *ellipsis size* reflects the fact that the resolution of *small ellipsis* is faster than the resolution of *large ellipsis* for both *also* conditions, while the main effect of the *position of also* is caused because the processing of the *high also* conditions is faster than the processing of the *low also* conditions. There were no other significant effects in the data. The results of the model for the third word after the auxiliary site are summarized in Table 5.

| Predictor | Coefficient | Standard Error | *t* value |
|---|---|---|---|
| Intercept | -29.634 | 9.056 | -3.272 |
| *Position of also* | 17.017 | 7.584 | 2.244 |
| *Ellipsis size* | 21.984 | 9.215 | 2.386 |
| *Position of also*×*Ellipsis size* | -6.791 | 11.810 | -0.575 |

Table 5: Results of Experiment 3b

Finally, an analysis was conducted pooling data from the *high also* condition across experiments 3a-b.[26] A linear mixed effects model predicting RRTs from the two factors of interest: *determiner* (*every* vs. *which*) and ellipsis size (*small ellipsis*, marked by *did*, vs. *large ellipsis*, marked by *was*) was fit to the data. The model included random slopes and intercepts for the factors of interest. The results show a main effect of *ellipsis size* two and three words after the auxiliary site, such that *small ellipsis* was processed more quickly than *large ellipsis* (log likelihood tests comparing a model with and without the predictor of interest, $p$'s$<0.05$). In addition, the results show a *determiner*×*ellipsis size* interaction two words after the auxiliary site, driven by the fact that the difference between the *every* conditions is greater than the *which* conditions across the two *ellipsis size* conditions. The results of the model for the second word after the auxiliary site are summarized in Table 6.

| Predictor | Coefficient | Standard Error | *t* value |
|---|---|---|---|
| Intercept | -25.08 | 14.32 | -1.751 |
| *Determiner* | 34.87 | 11.12 | 1.810 |
| *Ellipsis size* | 52.06 | 19.27 | 4.680 |
| *Determiner*×*Ellipsis size* | -30.06 | 15.20 | -1.978 |

Table 6: Results of comparison of *high* conditions in Exp3a-b

### 4.2.4 Discussion

Several effects can be observed in both Experiment 3a and Experiment 3b. First, we find a similar effect of the presence of *also* on participants' behavior when the word '*also*' was read. We can take this effect to indicate that our participants were paying attention to this experimental manipulation.

Additionally, we find a main effect of *ellipsis size* in both experiments, occurring on the second and third word after the ellipsis site. This effect is more pronounced on the second word following the auxiliary verb, as can be seen in Figure 9. The graph on the left shows RRTs for sentences with an object relative clause hosted by *every* (Experiment 3a) and the graph on the right shows RRTs for sentences with an object relative clause hosted by *which* (Experiment 3b). Note that here there is no difference between the two *also* conditions in the two experiments.

[26]We thank an anonymous reviewer and Anonymized for suggesting this analysis to us.

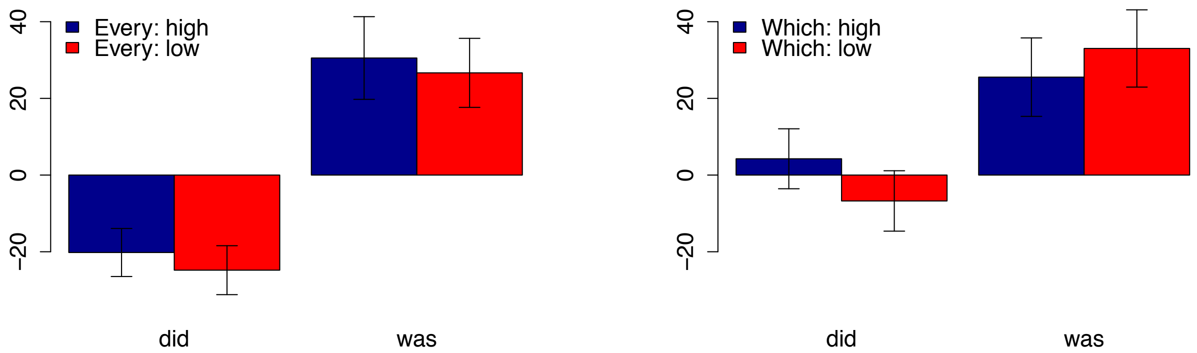


Figure 9: RRTs two words after AUX in Experiments 3a (*every*, left) and 3b (*which*, right)

In both Experiments 3a-b, we observe the main effect of *ellipsis size* not only on the second word following the auxiliary verb but also on the third word. More importantly, we find that on this word, the presence and position of *also* affects the resolution of ACD differently in the two experiments. Specifically, *also* does not have an effect in Experiment 3a, but in the case of Experiment 3b, we find a main effect of *position of also* in addition to the main effect of *ellipsis size*. To see this more closely, consider Figure 10, which compares the behavior of the four target conditions in Experiments 3a-b three words after the auxiliary verb. The graph on the left shows RRTs for sentences with an object relative clause hosted by *every* (Experiment 3a) and the graph on the right shows RRTs for sentences with an object relative clause hosted by *which* (Experiment 3b).

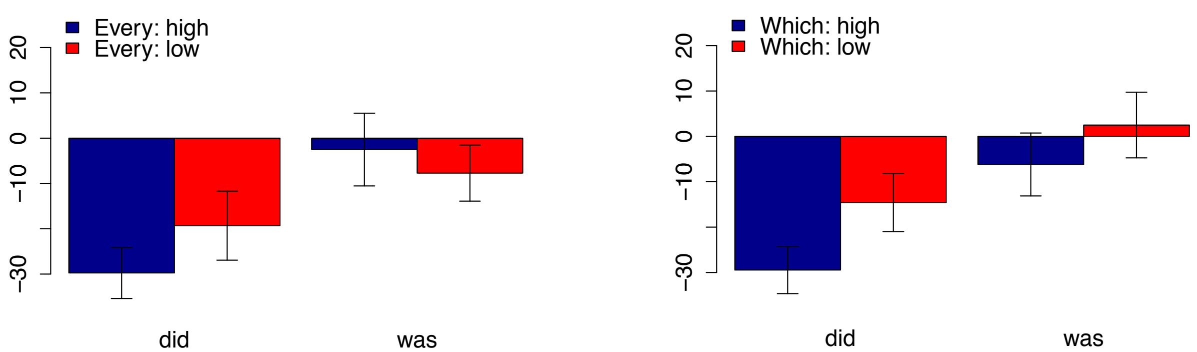


Figure 10: RRTs three words after AUX in Experiments 3a (*every*, left) and 3b (*which*, right)

The results of Experiment 3a exhibit parallel behavior for the two *also* conditions: the only effect here is that of the *ellipsis size*: *small ellipsis* (with *did*) is processed faster than *large ellipsis* (with *was*). In Experiment 3b, on the other hand, we find a main effect of *position of also* in addition to a main effect of *ellipsis size*. In particular, we find that the *high also* condition is processed faster than the *low also* condition across both *ellipsis* conditions.

The main effect of *position of also* in Experiment 3b, crucially exhibiting a facilitation effect of non-local ACD resolution with *high also* (i.e. of the *was-high also* condition) so that it is processed at a similar speed to that of the *local* ACD with *low also* (i.e. of the *did-low also* condition), is consistent with the partial movement approach. Specifically, if interpretability requirements force long-distance covert *wh*-movement, we expect ACD to be facilitated in the entire movement domain. In Experiment 3b, the *high also* forces *wh*-movement above it and as a result, antecedent containment is preemptively undone not only in the case of *small ellipsis* but also in the case of *large ellipsis*. When the parser reaches the gap site in both conditions, all it has to do is find an appropriate antecedent for the ellipsis and nothing more. This should translate into a facilitation effect both for the *small ellipsis* (marked with did) and for the *large ellipsis* (marked with was) for sentences in the *high also* condition.

Note that comparing the *high also* conditions across the two experiments yields an interaction, attributable to the fact that the difference between the *small* and *large* ellipsis conditions with *every* is greater than the difference between the two conditions with *which*. This, again, is predicted if *high also* facilitates the resolution of ellipsis in both the *small* and *large* conditions, but with *every* only *small* ellipsis resolution is facilitated. This effect can be observed in figure 11.[27]

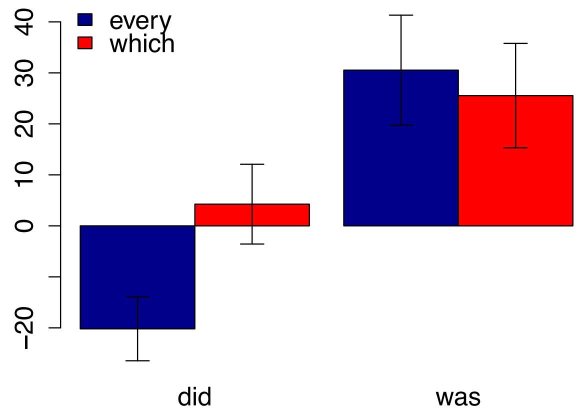



Figure 11: *Every* and *which* in Exp3

The effect of *ellipsis size* found in Experiments 3a-b is consistent with the effect found in the other experiments presented in the paper: resolving ellipsis with a larger, more complex antecedent, is more costly than resolving ellipsis with a smaller, simpler antecedent. Furthermore, this result shows up in the same region as in previous experiments: two and three words after the auxiliary verb. Here we see that this result is independent from the effect of *also*: although an intervener may force movement that preemptively undoes antecedent containment, resolving the ellipsis to a larger verbal complex is nonetheless more difficult than resolving it to a smaller, more local one. Because it is otherwise not possible to predict where the effect of *also* should occur, Experiments 3a-b were each replicated with the same items but different participants, to ensure that the results remain consistent. As noted in footnote 24, the data reported above reflects the collected data from these two separate runs of the experiment, and it is consistent with each individual set of results.

# 5 General discussion

The results of Experiments 1-3 shed light on the interaction between properties of in-situ *wh*-phrases and the resolution of Antecedent Contained Deletion. More specifically, we have argued that our findings suggest the following distribution at LF: (a) in-situ *wh*-phrases cannot be interpreted in their base position, but (b) they do not necessarily move to

[27] Recall that different participants contributed the data in the *which* vs. *every* conditions and therefore a direct comparison of RTs across *determiner* conditions is not possible here.

C for interpretation. Instead, we proposed the *partial movement* approach to *wh*-in-situ: in-situ *wh*-phrases can be interpreted at any propositional node in the structure, just like traditional quantifiers. Unlike traditional quantifiers, however, in-situ *wh*-phrases are subject to intervention effects. Thus, the presence of an element like *also*, which projects a domain of intervention, can force in-situ *wh*-phrases to move higher than a traditional quantifier in order to escape an intervention effect. We argued that this lends further support to the *partial movement* approach to *wh*-in-situ. Below we summarize the conclusions of our experiments, and then discuss one alternative interpretation of the results.

## 5.1 Evidence against an in-situ approach to *wh*-in-situ

The evidence that supports our claims came in the form of effects that different determiners have on Antecedent Contained Deletion resolution. To see how facilitation of ACD resolution might arise as a consequence of the position of an upstream determiner at LF, recall that ACD resolution involves at least three steps: (a) creating a structure in which antecedent containment is undone; (b) identifying an appropriate antecedent for the ellipsis; and (c) filling the antecedent into the gap and computing the resulting meaning. The latter two steps are required for all cases of VP ellipsis. The former movement step—the reanalysis of the structure so as to undo antecedent containment—is only required in the case of ACD.

We argue that the facilitation effects of ACD resolution that we find in our experiments are the result of an early movement step in the parse of the structure that preemptively undoes antecedent containment in that structure. As a result, under certain experimental manipulations, step (a) of ACD resolution need not be performed when the presence of an auxiliary verb, marking the presence of ellipsis in the structure, is detected by the parser because sufficient movement had already been assumed earlier in the parse. Such an early movement step may be triggered by quantifiers in object position, as shown in Hackl et al. (2012). Our experiments ask whether in-situ *wh*-phrases similarly undergo covert movement, and if so, what position this movement triggers.

Experiments 1-3 gave rise to the conclusion that *small* ellipsis (marked with *did*) is easier to process than *large* ellipsis (marked with *was*), regardless of the determiner heading the relative clause hosting the ellipsis (*which*, *every*, or *the*) . This is a consistent effect, showing up in the same place in all of our experiments. This result can be explained in a variety of ways, as discussed in more detail in section 2.4.3: it may be the case that the difficulty associated with the processing of *large* ellipsis stems from the longer covert movement step required for its resolution. Alternatively, it may be the case that a richer, more complex antecedent is more difficult to integrate into the meaning of the sentence, regardless of whether or not movement has happened. As a result, we have refrained from interpreting this result as indicating the presence of a covert movement step in our items.

Experiments 1-2 showed that *which* patterns with *every* in two ways: first, as noted above, *small* ellipsis is resolved faster than *large* ellipsis, with no difference in the amount of facilitation between these two determiners. Moreover, when the relative clause hosting the ellipsis is introduced by either *which* or *every*, it is easier to process than if it is introduced by *the*. We have argued that this finding is inconsistent with an in-situ approach to *wh*-in-situ. If *wh* were truly LF-in-situ, it should pattern with *the*—a non-quantificational determiner that does not require any movement for its interpretation—and not with *every*.

The fact that in Experiment 3 the *high also* condition is processed at a faster rate than *low also* with *which* but not with *every* provides an additional argument against the in-situ approach to *wh*-in-situ: under the covert movement approach to *wh*-in-situ we expect *also* to force movement above it to avoid an intervention effect, but no such movement is expected for a quantifier like *every*. However, if *wh*-phrases are truly LF-in-situ, there is no explanation for the facilitation effect of *high also* observed in Experiment 3b.

These results provide a new argument to help resolve a long-standing debate in the literature on the syntax and semantics of *wh*-questions: as outlined in the introduction, the literature adopts both the in-situ approach and the covert movement approach to the interpretation of surface in-situ *wh*-phrases. Despite decades of research, very few arguments have been given in favor of one view over the other (cf Pesetsky, 2000; Nissenbaum, 2000). This paper thus makes an important contribution to this debate: **surface in-situ *wh*-phrases in English questions must undergo a covert movement step for their interpretation; they do not occupy their base-generated position at LF, contra the predictions of the in-situ approach *wh*-in-situ.**[28]

## 5.2 Evidence for the *partial movement* approach to *wh*-in-situ

As we have seen, the results of Experiments 1-3 are incompatible with the in-situ approach to *wh*-in-situ. We have argued, moreover, that they are also incompatible with the common view of the covert movement approach which predicts that movement must target the interrogative complementizer, predicting non-local covert movement in our items, contrary to our findings. As described above, we found that *which* patterns with *every* in its ability to facilitate ACD resolution in Experiments 1-2: *every* and *which* were not different from one another in Experiment 1, and they were both faster than parallel items with the non-quantificational *the* in Experiment 2.

The facilitation behavior observed with *which* is explained if the parser must assume a small QR step of the DP hosting the object relative clause to the nearest propositional node in order to integrate it into the structure. This is similar to the arguments made in Hackl et al. (2012) for *every*: a quantifier in object position cannot semantically compose with the verb in-situ, and requires a QR step to the nearest propositional node for interpretation. This early QR step preemptively undoes antecedent containment in the case of *small* ellipsis: step (a) of ACD resolution can be avoided in this case, resulting in faster reading

[28]More cautiously, this finding may only be accurate for *superiority-obeying* questions in English. Pesetsky (2000) shows that there may be reason to treat superiority-violating questions differently than superiority-obeying ones. Testing superiority-violating questions is beyond the scope of the current paper.

times. This step cannot be avoided for the *large* ellipsis, leading to higher processing costs. For *the*, no covert movement is assumed when it is encountered by the parser, because *the* can be composed as the sister of a verb in-situ. The parser assumes covert movement in the structure only when it reaches the gap site of ACD, resulting in slower reading times for both *small* and *large* ellipsis with *the*.[29]

To explain this behavior, we proposed the *partial movement* approach to *wh*-in-situ. This approach can be seen as an amendment to the covert movement approach, which draws on insights from the in-situ approach: ***wh*-phrases must move for basic structure-building reasons and interpretability, but they need not move further than is required by these external considerations. Once a *wh*-phrase is in an interpretable position, it can make its contribution to the interrogative semantics without moving to C.** As a consequence, we require a theory of interrogative syntax/semantics that allows in-situ *wh*-phrases to be interpreted in positions other than C, and immediately below C as well.[30]

Experiment 3 corroborated an important prediction of the *partial movement* approach: that it should be possible to distinguish the behavior of *which* from that of *every*, and in particular that if a *wh*-phrase is forced to move even higher than its QRed position for interpretation, we should find facilitation effects in the entire domain of movement. Experiment 3 tested this prediction by using intervention effects, which have been argued to require movement of *wh*-phrases above an intervener for interpretation, but do not affect the behavior of other quantifiers such as *every*. We argued that *also* is an intervener in English and showed that, indeed, when *also* forces non-local movement of an in-situ *wh*-phrase, we see facilitation effects in ACD resolution of local, and crucially also of non-local ACD. The sensitivity of *which*, and the insensitivity of *every*, to the same manipulation show that *also* does not simply act as a scope marker for covert movement: such a proposal would incorrectly predict similar effects on *every* and *which*. Instead, we see *also* affecting only the *wh*-phrase but not the quantifier, consistent with the predictions of the intervention effect theory (Beck 2006) and the predictions of the *partial movement* approach.[31]

If the *partial movement* approach to *wh*-in-situ is on the right track, traditional approaches to interrogative syntax/semantics cannot be correct. *Wh*-phrases cannot be interpreted where they are pronounced, as in-situ approaches would have it, but they also need not move to C for their interpretation, as covert movement approaches traditionally claim. In-

[29]These results are consistent with the results of Xiang et al. (2014), who compare the processing of in-situ *wh*-phrases to non-*wh* counterparts in Mandarin Chinese using a Speed-Accuracy Tradeoff experiment. As Mandarin is a *wh*-in-situ language, this study was able to compare simplex questions to minimally different sentences that did not contain a *wh*-phrase. Xiang et al. (2014) observe a processing cost involved with the interpretation of in-situ *wh*-phrases which they argue may be attributable to covert movement. However, we note that Xiang et al. (2014) do not rule out the possibility of alternative mechanism for the interpretation of *wh*-phrases in their experiment, which do not involve movement. These mechanisms will not be able to explain the data presented here, as discussed above.

[30]The development of such a theory goes beyond the scope of the present paper. See Cable (2007, 2010) and Kotek (2014a) for proposals that contains the ingredients to interpret structures like we propose.

[31]We note that as a consequence, any theory of intervention effects that derives intervention without any covert movement faces difficulties in explaining the results of Experiment 3.

stead, *wh*-phrases may remain in their lowest interpretable position, and from there make their contribution to interrogative semantics without any further movement.

## 5.3 An alternative proposal

Before concluding this paper, we would like to address one alternative interpretation of the experimental results presented above. This alternative interpretation will be compatible with our finding that the in-situ approach to *wh*-in-situ cannot be correct, but it attempts to reconcile our findings with the traditional covert movement approach, under which covert *wh*-movement necessarily targets interrogative C.[32]

Let us return to our sample item from Experiment 1, repeated here for convenience. Recall that this experiment compared the behavior of *every* with that of *which* with regard to facilitation of ACD resolution, and found that they facilitate ACD resolution to the same extent. We repeat the graph showing the results for two words after the auxiliary verb.

(35) **Sample target item in Experiment 1**

The conductor asked *which soloist* was [$_{VP1}$ willing to [$_{VP2}$ perform...

a. **every** concerto that the brilliant protégé <u>*did*</u> ... *(small ellipsis)*

b. **which** concerto that the brilliant protégé <u>*did*</u> ... *(small ellipsis)*

c. **every** concerto that the brilliant protégé <u>*was*</u> ... *(large ellipsis)*

d. **which** concerto that the brilliant protégé <u>*was*</u> ... *(large ellipsis)*

...and restructured the rehearsal accordingly.

Our starting point here was the assumption that *every* undergoes a short QR movement step to a position above the lower $VP_2$—here, *perform*—in order to resolve a type-mismatch as soon as it is encountered by the parser. This assumption was based on the results of Hackl et al. (2012), surveyed above in section 2.3, who showed that such movement indeed takes place in their items. Note, however, that our items differ from Hackl et al.'s (2012) in that they are embedded inside a question.

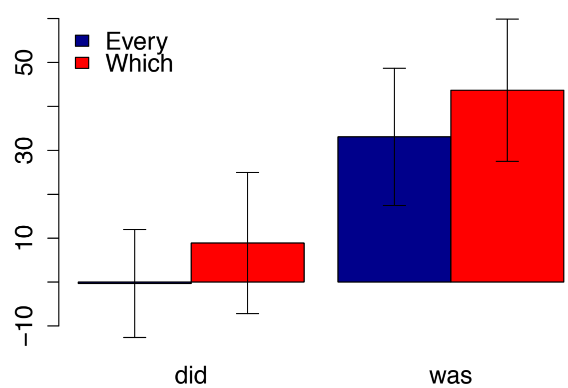



Figure 12: Results of Experiment 1

Questions with quantifiers are known to be ambiguous between two readings:

[32]We would like to thank Anonymized for suggesting this idea and discussing its implications with us.

(36) **The two readings of a question with a quantifier**
*Which* book did *every boy* read?
a. Which book is such that all the boys read it? *which* > ∀
b. For each boy, which book did he read? ∀ > *which*

The reading in (36a) involves narrow scope for *every* with respect to *which*, and is consistent with our assumptions that *every* only undergoes a short movement step. On the other hand, reading in (36b) involves wide scope for *every* with respect to *which*, and would require *every* to take scope above the question. There are several theories of how this scope is obtained, but it is at least possible that it is obtained through long-distance covert movement of *every* (cf Chierchia, 1993; Krifka, 2006; Nicolae, 2013).

Our methodology does not allow us to test which interpretation was accessed by our speakers. As a result, we must entertain the idea that at least some speakers accessed this reading for some of the items. This is made more plausible by the fact that multiple questions themselves admit a similar ambiguity, between a single-pair and a pair-list reading (Dayal, 2002, a.o.). At least the pair-list reading, and perhaps also the single-pair, have been argued to require covert movement for their interpretation (Dayal, 2002; Kotek, 2014a)

(37) **The two readings of a multiple question**
*Which* book did *which* boy read?
a. Which book and which boy are such that the boy read the book? *single-pair*
b. For each book, which boy read it? *pair-list*

If our items are read with a pair-list reading, then we expect *every* to undergo long-distance movement to a position above the question—and hence above both the embedded $VP_2$ and the higher $VP_1$. Consequently, ACD should be preemptively undone in both the *small* and *large* ellipsis conditions, and ACD resolution should be facilitated when the parser encounters both <u>*did*</u> and <u>*was*</u>. The fact that *every* patterns with *which*, then, teaches us not that *every* and *which* both underwent a short QR movement step just above $VP_2$, but instead that they both underwent long-distance movement, targeting a position higher than $VP_1$. The difference between the *small* and *large* ellipsis could be attributed to the fact that the larger ellipsis is richer and therefore may be more costly to integrate into the structure than a simpler antecedent, and not to any additional covert movement in the structure.

The results of Experiment 2 could be consistent with this view of interrogative syntax: we observe that *every* and *which* are both processed faster than corresponding sentences with *the*, and this could be explain both under a theory where both have undergone a short movement step (as we proposed in the previous section) or with the theory entertained here, that they both underwent a long-distance covert movement step.

However, we believe that this alternative proposal is inconsistent with the results of Experiment 3. If in-situ *wh*-phrases always covertly move to C, then *also* should not be able to affect *which*, since sufficient movement to avoid an intervention effect would be performed regardless of the presence and position of *also*. Consequently, we may propose that instead, *also* interacts with *every* so as to slow down the processing of the items in Experiment 3a. One possible proposal is that *also* marks the possible scope of *every*: *low also* would force a short QR step, and *high also* would be compatible with a longer movement step. Crucially, we would have to assume that such a constraint does not affect the movement of *which*, as the covert movement approach to *wh*-in-situ requires that *which* always target interrogative C. However, if this were the case, we would expect an interaction in Experiment 3a, such that the *low also-was* condition suffers, as it would be the only condition in that experiment where movement of *every* is unable to preemptively undo antecedent containment. This is not the result observed in Experiment 3a: we instead observe a main effect of ellipsis size but no interaction. This result was replicated twice, as described in section 4.2, to ensure its validity.

To summarize, it seems that the proposal that questions with *every* are assigned a pair-list reading, consistent with a long-distance step of the quantifier, may be true in some cases, but a proposal that attempts to explain our data based solely on pair-list LFs in which *every* and *which* need to take non-local scope is not consistent with the results of Experiment 3. Only if an alternative proposal is put forth that can explain the results of this experiment can we reconcile the results of our experiments with the traditional covert movement approach, which requires *wh*-phrases to always move to C for interpretation. At the moment we are unaware of such an proposal. As a consequence, we propose to adopt the *partial movement* approach to *wh*-in-situ, allowing in-situ *wh*-phrases to be interpreted not only at C but at other nodes in the structure as well.[33] Regardless of how this debate is settled, we have shown that the in-situ approach to *wh*-in-situ cannot be correct for English questions, thus contributing to the debate over the architecture of questions.

# 6 Conclusion

We have provided evidence for two strong claims:

(38) **Conclusions from Experiments 1-3**

a. *Wh*-in-situ in English multiple questions always undergo at least some covert movement in the course of the derivation of a question.

b. *Wh*-in-situ may move to any node with propositional type at LF. All things being equal, movement will be the shortest that will yield an interpretable structure. Movement need not target interrogative C.

These conclusions resolve the longstanding debate regarding the nature of *wh*-in-situ in a novel way. Traditional approaches propose that (a) *wh* moves to interrogative C for

[33] See also recent offline evidence that such a proposal may be necessary for English superiority-obeying questions in Kotek (2014b).

interpretation (the covert movement approach), or alternatively (b) *wh* stays in its base-generated position and is interpreted without movement (the in-situ approach). We have shown that a combined approach fares better at explaining the experimental findings described in this paper: *wh* does not stay in-situ: it must always move at least a little bit. However, *wh* can be interpreted at any node with propositional type in the structure, and thus movement is not driven by a need to reach C but simply by a need to reach an interpretable position. Both findings constrain the theories of that can correctly describe English questions, and as a consequence contribute to our understanding of the architecture of interrogative syntax/semantics.

# References


Anderson, Catherine. 2004. The structure and real-time comprehension of quantifier scope ambiguity. Doctoral Dissertation, Northwestern University.

Aoun, Joseph, and Yen-hui Audrey Li. 1993. *Wh*-elements in situ: Syntax or LF? *Linguistic Inquiry* 24:199–238.

Arnon, Inbal, Neal Snider, Philip Hofmeister, Florian T. Jaeger, and Ivan A. Sag. To appear. Cross-linguistic variation in a processing account: The case of multiple *wh*-questions. In *Proceedings of BLS 32*.

Arregui, Ana, Charles Jr. Clifton, Lyn Frazier, and Keir Moulton. 2006. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language* 55:232–246.

Baayen, R. H. 2004. Statistics in psycholinguistics: A critique of some current gold standards. In *Mental Lexicon Working Papers*, 1–45. Edmonton.

Baker, Carol L. 1970. Notes on the description of English questions: the role of an abstract question morpheme. *Foundations of Language* 6:197–217.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry Tily. 2013. Random effects structure for confirm-atory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68:225–278.

Bates, Douglas M., and Deepayan Sarkar. 2007. lme4: Linear mixed-effects models using S4 classes. R package version 0.9975-12, URL http://CRAN.R-project.org/.

Beaver, David, and Brady Clark. 2008. *Sense and sensitivity: How focus determines meaning*. Wiley-Blackwell.

Beck, Sigrid. 1996. Quantified structures as barriers for LF movement. *Natural Language Semantics* 4:1–56.

Beck, Sigrid. 2006. Intervention effects follow from focus interpretation. *Natural Language Semantics* 14:1–56.

Beck, Sigrid, and Shin-Sook Kim. 1997. On *wh*-and operator scope in Korean. *Journal of East Asian Linguistics* 6:339–384.

Bever, Tom. 1970. The cognitive basis for linguistic structure. In *Cognition and the development of language*, ed. J. R. Hayes, 279–362. New York: Wiley.

Cable, Seth. 2007. The grammar of Q. Doctoral Dissertation, Massachusetts Institute of Technology.

Cable, Seth. 2010. *The grammar of Q: Q-particles, wh-movement, and pied-piping*. Oxford University Press.

Cheng, Lisa Lai-Shen. 1991. On the typology of *wh*-questions. Doctoral Dissertation, Massachusetts Institute of Technology.

Chierchia, Gennaro. 1993. Questions with quantifiers. *Natural Language Semantics* 1:181–234.

Chomsky, Noam. 1981. *Lectures on government and binding*. Foris.

Chomsky, Noam. 1991. Some notes on economy of derivation and representation. In *Principles and parameters in comparative grammar*. MIT Press.

Chomsky, Noam. 1993. A minimalist program for linguistic inquiry. In *The view from Building 20*, ed. Kenneth Hale and Samuel Jay Keyser. MIT Press.

Chomsky, Noam. 1995. *The minimalist program*. MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: the framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*. MIT Press.

Clifton, Charles Jr., Gisbert Fanselow, and Lyn Frazier. 2006. Amnestying superiority violations: processing multiple questions. *Linguistic Inquiry* 37:51–68.

Clifton, Charles Jr., Lyn Frazier, and Patricia Deevy. 1999. Feature manipulation in sentence comprehentsion. *Rivista di Linguistica* 11:11–39.

Collins, Chris. 2001. Economy conditions in syntax. In *Handbook of syntactic theory*, ed. Chris Collins and Mark R. Baltin. Blackwell publishing.

Dayal, Veneeta. 2002. Single-pair versus multiple-pair answers: *wh*-in-situ and scope. *Linguistic Inquiry* 33:512–520.

Dornisch, Ewa. 2000. Overt quantifier raising in Polish. In *Proceedings of Generative Linguistics in Poland GLiP*, volume 1, 47–58.

van Dyke, Julie A., and Richard L. Lewis. 2003. Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language* 49:285–316.

van Dyke, Julie A., and Brian McElree. 2006. Retrieval interference in sentence comprehension. *Journal of Memory and Language* 55:157–166.

É Kiss, Katalin. 1986. Against the LF-movement of *wh*-phrases. Ms., Hungarian Academy of Sciences, Budapest.

Epstein, Sam. 1992. Derivational constraints on A' chain formation. *Linguistic Inquiry* 23:235–259.

Fanselow, Gisbert, Denisa Lenertova, and Thomas Weskott. 2008. Studies on the acceptability of object movement to Spec,CP. In *Language content & cognition: The discourse potential of underspecified structures*, ed. Anita Steube. Mouton de Gruyter.

Featherston, Sam. 2005a. Magnitude estimation and what it can do for your syntax: some *wh*-constraints in German. *Lingua* 115:1525–1550.

Featherston, Sam. 2005b. Universals and grammaticality: *Wh*-constraints in German and English. *Linguistics* 43:667–711.

Foraker, Stephani, and Brian McElree. 2011. Comprehension of linguistic dependencies: Speed-accuracy tradeoff evidence for direct-access retrieval from memory. *Language and Linguistics Compass* 5:764–783.

Ford, Marylin, Joan Bresnan, and Ron Kaplan. 1982. A competence-based theory of syntactic closure. In *The mental representation of grammatical relations*, ed. Joan Bresnan, 727–796. Cambridge, Mass.: MIT Press.

Fox, Danny. 1995. Economy and scope. *Natural Language Semantics* 3:283–341.

Fox, Danny. 2000. *Economy and semantic interpretation*. Cambridge, Mass.: MIT Press.

Fox, Danny. 2002. Antecedent-contained deletion and the copy theory of movement. *Linguistic Inquiry* 33:63–96.

Fox, Danny. 2003. *On Logical Form*, chapter 2, 82–123. Oxford, UK: Blackwell Publishers.

Frazier, Lyn. 1999. *On sentence interpretation*. Dordrecht: Kluwer.

Frazier, Lyn, and Charles Jr. Clifton. 2000. On bound variable interpretations: The LF-only hypothesis. *Journal of Psycholinguistic Research* 29:125–139.

Frazier, Lyn, and Charles Jr. Clifton. 2001. Parsing coordinates and ellipsis: Copy $\alpha$. *Syntax* 4:1–22.

Frazier, Lyn, and Charles Jr. Clifton. 2002. Processing D-linked phrases. *Journal of Psycholinguistic Research* 31:633–660.

Frazier, Lyn, and D. Fodor, Janet. 1978. The sausage machine: A new two-stage parsing model. *Cognition* 6:291–325.

Frazier, Lyn, and Keith Rayner. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology* 14:178–210.

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68:1–76.

Gordon, Peter C., Randall Hendrick, and Marcus Johnson. 2004. Effects of noun phrase type on sentence complexity. *Journal of Memory and Language* 51:97–114.

Hackl, Martin, Jorie Koster-Hale, and Jason Varvoutis. 2012. Quantification and ACD: Evidence from real-time sentence processing. *Journal of Semantics* 29:145–206.

Hagstrom, Paul. 1998. Decomposing questions. Doctoral Dissertation, Massachusetts Institute of Technology.

Hamblin, Charles. 1973. Questions in Montague English. *Foundations of Language* 10:41–53.

Heim, Irene, and Angelika Kratzer. 1998. *Semantics in generative grammar*. Blackwell.

Hofmeister, Philip, Florian T. Jaeger, Ivan A. Sag, Inbal Arnon, and Neal Snider. 2007. Locality and accessibility in *wh*-questions. In *Roots: Linguistics in search of its evidential base*, ed. Sam Featherston and Wolfgang Sternfeld. Mouton de Gruyter.

Hoji, Hajime. 1985. Logical Form constraints and configurational structures in Japanese. Doctoral Dissertation, University of Washington.

Hornstein, Norbert. 1995. *Logical Form: from GB to Minimalism*. Cambridge, Mass.: Blackwell Publishers.

Huang, C. T. James. 1995. Logical Form. In *Government and binding theory and the minimalist program*, ed. G. Webelhuth, 127–173. Blackwell.

Huang, Cheng-Teh James. 1982. Move *wh* in a language without *wh* movement. *The Linguistic Review* 1.

Johnson, Kyle. 1996. When verb phrases go missing. *Glot International* 2.5:3–9.

Just, Marcel A., Patricia A. Carpenter, and Jacqueline D. Wooley. 1982. Paradigms and processes and in reading comprehension. *Journal of Experimental Psychology: General* 111:228–238.

Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1:3–44.

Kazanina, Nina, and Colin Phillips. 2000. Differential effects of constraints in the processing of russian cataphora. *The Quarterly Journal of Experimental Psychology* 63:371–400.

Kehler, Andrew. 2001. *Coherence, reference, and the theory of grammar. center for the study of language and information*. CSLI Publications.

Kertz, Laura. 2010. Ellipsis reconsidered. Doctoral Dissertation, University of California Davis.

Kim, Shin-Sook. 2002. Focus matters: Two types of intervention effects. In *Proceedings of WCCFL 21*.

Kim, Soowon. 1991. Chain scope and quantification structure. Doctoral Dissertation, Brandeis University.

Kishimoto, Hideki. 2005. *Wh*-in-situ and movement in Sinhala questions. *Natural Language & Linguistic Theory* 23:1–51.

Kitahara, Hisatsugu. 1997. *Elementary operations and optimal derivations*. Cambridge, Mass.: MIT Press.

Kotek, Hadas. 2014a. Composing questions. Doctoral Dissertation, Massachusetts Institute of Technology.

Kotek, Hadas. 2014b. Intervention out of islands. In *Proceedings of NELS 44*, ed. Leland Kusmer and Jyoti Iyer, volume 1, 234—246. Amherst: GLSA.

Kratzer, Angelika, and Junko Shimoyama. 2002. Indeterminate pronouns: the view from Japanese. In *The Proceedings of the Third Tokyo Conference on Psycholinguistics (TCP 2002)*.

Krifka, Manfred. 2006. Association with focus phrases. In *The architecture of focus*, 105–136. Mouton de Gruyter.

Larson, Richard K., and Robert Carlen May. 1990. Antecedent containment or vacuous movement: Reply to Baltin. *Linguistic Inquiry* 21:103–122.

Lasnik, Howard, and Mamoru Saito. 1992. *Move $\alpha$, conditions on its application and output*. Cambridge, Mass.: MIT Press.

Lewis, Richard L., Shravan Vasishth, and Julie A. van Dyke. 2006. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences* 10:447–454.

MacDonald, Maryellen C., Neal J. Perlmutter, and Mark S. Seidenberg. 1994. Lexical nature of syntactic ambiguity resolution. *Psychological review* 101:676–703.

Martin, Andrea E., and Brian McElree. 2008. A content-addressable pointer underlies comprehension of verb–phrase ellipsis. *Journal of Memory and Language* 58:879–906.

Martin, Andrea E., and Brian McElree. 2009. Memory operations that support language comprehension: Evidence from verb–phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35:1231–1239.

Martin, Andrea E., and Brian McElree. 2011. Direct-access retrieval during sentence comprehension: Evidence from sluicing. *Journal of Memory and Language* 64:327–343.

May, Robert Carlen. 1985. *Logical Form: Its structure and derivation*. Linguistic Inquiry Monographs. MIT Press.

Mayr, Clemens. to appear. Intervention effects and additivity. *Journal of Semantics* .

Merchant, Jason. 2013. Voice and ellipsis. *Linguistic Inquiry* 44:77–108.

Murphy, Gregory. 1985. Processes of understanding anaphora. *Journal of Memory and Language* 24:290–303.

Nicolae, Andreea. 2013. Any questions? Doctoral Dissertation, Harvard.

Nishigauchi, Taisuke. 1986. Quantification in syntax. Doctoral Dissertation, University of Massachusetts Amherst.

Nissenbaum, Jon. 2000. Covert movement and parasitic gaps. In *Proceedings of NELS 30*, ed. M. Hirotani, Coetzee A., Hall N., and Kim J.-Y., volume 2, 541–555.

Pesetsky, David. 1987. *Wh*-in-situ: movement and unselective binding. In *The representation of (in)definiteness*. MIT Press.

Pesetsky, David. 2000. *Phrasal movement and its kin*. Cambridge, Mass.: MIT Press.

Phillips, Colin. 1996. Order and structure. Doctoral Dissertation, Massachusetts Institute of Technology.

Phillips, Colin. 2003. Linear order and constituency. *Linguistic Inquiry* 34:37–90.

Polinsky, Maria, Carlos Gomez Gallo, Peter Graff, and Ekaterina Kravtchenko. 2012. Subject preference and ergativity. *Lingua* 122:267–277.

Polinsky, Maria, and Eric Potsdam. Forthcoming. Left edge topics in Russian and the processing of anaphoric dependencies. *Journal of Linguistics* .

Pritchett, Bradley L. 1992. *Grammatical competence and parsing performance*. The University of Chicago Press.

Reinhart, Tanya. 1998. *Wh*-in-situ in the framework of the Minimalist Program. *Natural Language Semantics* 6:29–56.

Reinhart, Tanya. 2006. *Interface strategies: Optimal and costly computations*. Cambridge, Mass.: MIT Press.

Richards, Norvin. 2001. *Movement in language: Interactions and architectures*. Oxford University Press.

Rullmann, Hotze, and Sigrid Beck. 1998. Reconstruction and the interpretation of *which*-phrases. In *Proceedings of the 1997 Tübingen Workshop*, ed. G. Katz, S.-S Kim, and H. Winhart.

Sag, Ivan Andrew. 1976. Deletion and Logical Form. Doctoral Dissertation, Massachusetts Institute of Technology.

San Pietro, Steven A., Ming Xiang, and Jason Merchant. 2012. Accounting for voice mismatches in ellipsis. In *Proceedings of WCCFL 30*, ed. Nathan Arnett and Ryan Bennett, 303–312. Somerville, MA: Cascadilla Proceedings Project.

Shimoyama, Junko. 2006. Indeterminate quantification in Japanese. *Natural Language Semantics* 14:139–173.

Soh, Hooi Ling. 2005. *Wh*-in-situ in Mandarin Chinese. *Linguistic Inquiry* 36:143–155.

Tanenhaus, Michael K., and John C. Trueswell. 1995. Sentence comprehension. In *Speech, language, and communication. handbook of perception and cognition*, ed. Joanne L. Miller and Peter D. Eimas, volume 11, 217–262. San Diego, CA, US: Academic Press, 2nd edition.

Tomioka, Satoshi. 2007a. Intervention effects in focus: From a Japanese point of view. In *Proceedings of the 2nd Workshop on Prosody, Syntax, and Information Structure (WPSI 2)*, ed. Shinichiro Ishihara, 97–117. Potsdam: Potsdam University Press.

Tomioka, Satoshi. 2007b. Pragmatics of LF intervention effects: Japanese and Korean interrogatives. *Journal of Pragmatics* 39:1570–1590.

Tsai, Wei-tien Dylan. 1994. On nominal islands and LF extractions in Chinese. *Natural Language and Linguistic Theory* 12:121–175.

Tunstall, Susanne L. 1998. The interpretation of quantifiers: Semantics and processing. Doctoral Dissertation, University of Massachusetts Amherst.

Wagers, Matthew W., and Colin Phillips. 2009. Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics* 45:395–433.

Warren, Tessa, and Edward Gibson. 2002. The influence of referential processing on sentence complexity. *Cognition* 85:79–112.

Watanabe, Akira. 2001. *Wh*-in-situ languages. In *The handbook of contemporary syntactic theory*, ed. Mark R. Baltin and Chris Collins, 203–225. Oxford, UK: Blackwell.

Williams, Edwin. 1974. Rule ordering in grammar. Doctoral Dissertation, Massachusetts Institute of Technology.

Williams, Edwin. 1977. Discourse and Logical Form. *Linguistic Inquiry* 8:101–139.

Xiang, Ming, Brian Dillon, Matthew W. Wagers, Fengqin Liu, and Taomei Guo. 2014. Processing covert dependencies: an sat study on mandarin *wh*-in-situ questions. *Journal of East Asian Linguistics* 23:207–232.

Xiang, Ming, Boris Harizanov, Maria Polinsky, and Ekaterina Kravtchenko. 2011. Processing morphological ambiguity: An experimental investigation of russian numerical phrases. *Lingua* 121:548–560.

# Appendix A: Materials for Experiments 1-3

Below is the full of items used in Experiments 1-3. The same item paradigms were used in all three experiments, with different choices of determiner and placement/position of *also* as described in the body of the paper. In short: Experiments 1 and 3 compared the determiners *every* and *which*, while Experiment 2 compared the determiners *the* and *every*. Experiment 3 compared the behavior of a *high* and a *low* placement of *also*, while Experiments 1-2 did not contain *also*. Recall that in all experiments, the sentences were presented on two lines, with the line break always placed immediately following the verbal complex.

1. The orderly learned which doctor was (also) planning to (also) monitor *the/every/which* patient that the duty nurse *did*/*was* and immediately updated the charts.
2. The principal determined which instructor was (also) able to (also) teach *the/every/which* class that the substitute teacher *did*/*was* and accordingly finalized the schedule.
3. The conductor asked which soloist was (also) willing to (also) play *the/every/which* piece that the brilliant protégé *did*/*was* and restructured the rehearsal accordingly.
4. The coordinator learned which tutor was (also) scheduled to (also) teach *the/every/which* topic that the Physics professor *did*/*was* and assigned them to classrooms.
5. The prosecutor asked which witness was (also) told to (also) discredit *the/every/which* defendant that the corrupt detective *did*/*was* but only one witness revealed anything.
6. The teacher found_out which student was (also) eager to (also) attend *the/every/which* trip that the class president *did*/*was* and organized the field trips accordingly.
7. The detective found_out which guard was (also) willing to (also) hassle *the/every/which* prisoner that the sadistic warden *did*/*was* and included the names in his report.
8. The analyst predicted which investor was (also) prepared to (also) buy *the/every/which* stock that the hedge fund *did*/*was* and then sent a$_{m}$emo to the bank management.
9. The realtor asked which trainee was able to (also) show *the/every/which* property that the experienced secretary *did*/*was* but nobody was (also) available that weekend.
10. The carpenter asked which apprentice was (also) qualified to (also) use *the/every/which* technique that the licensed electrician *did*/*was* and then assigned personnel to projects.
11. The choreographer determined which dancer was (also) ready to (also) perform *the/every/which* dance_routine that the Russian ballerina *did*/*was* and then started the dance recital.
12. The organizers found_out which announcer was (also) willing to (also) cover *the/every/which* game that the notorious commentator *did*/*was* and then finalized the broadcasting schedule.

13. The librarian learned which teacher was (also) planning to (also) borrow *the/every/which* book that the visiting scholar *did*/*was* and accordingly shortened the loan periods.

14. The attorney clarified which witness was (also) supposed to (also) support *the/every/which* alibi that the undercover informant *did*/*was* and then gave his closing argument.

15. The dispatcher clarified which apprentice was (also) scheduled to (also) accompany *the/every/which* crew that the experienced engineer *did*/*was* and sent the crews on their way.

16. The programmer realized which update was (also) certain to (also) solve *the/every/which* problem that the old software *did*/*was* but surprisingly decided not to tell anyone.

17. The focus-group explained which discount was (also) likely to (also) attract *the/every/which* demographic that the Spring sale *did*/*was* and then several TV ads were launched.

18. The secretary found_out which professor was (also) going to (also) question *the/every/which* student that the disciplinary committee *did*/*was* and then scheduled the hearings.

19. The general forgot which unit was (also) scheduled to (also) attack *the/every/which* target that the nuclear submarine *did*/*was* and sent a messenger to headquarters.

20. The biologist discovered which reptile was (also) likely to (also) have *the/every/which* gene that the Tyrannosaurus Rex *did*/*was* and proposed additional tests.

21. The admiral specified which ship was (also) ordered to (also) attack *the/every/which* position that the navy jet *did*/*was* and then the joint army-navy exercise began.

22. The engineer explained which apprentice was (also) asked to (also) service *the/every/which* engine that the sick crew_member *did*/*was* and then called the train company.

23. The colonel explained which officer was (also) ordered to (also) interrogate *the/every/which* prisoner that the CIA agent *did*/*was* and then described what methods not to use.

24. The log showed which detective was (also) sent to (also) arrest *the/every/which* suspect that the FBI agent *did*/*was* and additionally where the arrest took place.

25. The detective discovered which mobster was (also) about to (also) blackmail *the/every/which* business that the street gang *did*/*was* and immediately informed his superiors.

26. The sheriff knew which marshal was (also) excited to (also) chase *the/every/which* fugitive that the state police *did*/*was* but doubted that the fugitives would be caught.

27. The scientist discovered which antibody was (also) likely to (also) attack *the/every/which* virus that the standard medication *did*/*was* but needed funding to complete her study.

28. The warden guessed which inmate was (also) trying to (also) smuggle *the*/*every*/*which* contraband that the corrupt guard *did*/*was* and therefore intensified the security screens.

## Appendix B: Offline ratings of the items in Experiments 1-3

Below is a description of offline ratings collected for the target sentences in all five experiments reported in this paper.[34] Each experiment was posted in full, using the same target and filler items as used in the self-paced reading experiments. Each experimental item was followed by a comprehension question, as in the corresponding self-paced reading experiment, and additionally by a naturalness / acceptability rating. 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.

Participants for these studies were recruited through Amazon Mechanical Turk and the studies themselves were hosted on Ibex Farm. 60 subjects participated in each study and they were paid at a rate of $3.5 an hour for their participation. The experiments took on average 30 minutes to complete. 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. Participants were excluded from the analysis if they participated in more than one study, if they indicated that their native language was not English, or if their accuracy on comprehension questions across the entire experiment was below 75%. In total, 7 participants were excluded across the five experiments: 3 in Experiment 1, 2 in Experiment 2a, and 1 in Experiment 3a.

In what follows, we give a summary of raw ratings for each experiment and describe the statistical significance of the results, and then provide a short discussion. In all cases, a linear mixed effects model with random slopes and intercepts was fit to the data, predicting z-transformed acceptability ratings (means and standard deviations estimated within participants) from the two factors of interest in each experiment (*ellipsis size* and *determiner* in Experiments 1-2a-b, and *ellipsis size* and *position of also* in Experiment 3a-b).

[34]We would like to thank Anonymized for help running these experiments, and Anonymized for technical help with the analysis of the results.

## Experiment 1

The raw ratings for Experiment 1 are given in table 7. Recall that this experiment compares the behavior of *every* and *which* across two *ellipsis size* conditions: *small ellipsis* (*did*) and *large ellipsis* (*was*).

| Condition | Rating |
|---|---|
| every-did | 3.16 |
| every-was | 2.65 |
| which-did | 2.61 |
| which-was | 2.55 |

Table 7: Ratings for target items in Experiment 1 (*every-which, did-was*)

The results for this experiment show a significant interaction ($p < 0.05$), such that the *every-did* condition is rated higher than the other conditions.

## Experiment 2a

The raw ratings for Experiment 2a are given in table 8. Recall that this experiment compares the behavior of *every* and *the* across two *ellipsis size* conditions: *small ellipsis* (*did*) and *large ellipsis* (*was*).

| Condition | Rating |
|---|---|
| every-did | 3.31 |
| every-was | 2.92 |
| the-did | 3.11 |
| the-was | 2.80 |

Table 8: Ratings for target items in Experiment 2a (*every-the, did-was*)

The results show a main effect of *ellipsis size* ($p < 0.05$), such that *small ellipsis* (with *did*) was rated higher than *large ellipsis* (with *was*) across both determiners. There were no other significant effects in the data.

## Experiment 2b

The raw ratings for Experiment 2b are given in table 9. Recall that this experiment compares the behavior of *which* and *the* across two *ellipsis size* conditions: *small ellipsis* (*did*) and *large ellipsis* (*was*).

| Condition | Rating |
|---|---|
| which-did | 2.82 |
| which-was | 2.46 |
| the-did | 2.98 |
| the-was | 2.66 |

Table 9: Ratings for target items in Experiment 2b (*which-the, did-was*)

The results show a main effect of *ellipsis size* ($p < 0.05$), such that *small ellipsis* (with did) was rated higher than *large ellipsis* (with was) across both determiners. There were no other significant effects in the data.[35]

## Experiment 3a

The raw ratings for Experiment 3a are given in table 10. Recall that this experiment compares the behavior of *every* across two *ellipsis size* conditions: *small ellipsis* (*did*) and *large ellipsis* (*was*), and two positions for the potential intervener *also*: *low also*, above the embedded predicate, and *high also*, above the matrix predicate.

| Condition | Rating |
|---|---|
| low-did | 3.32 |
| low-was | 3.09 |
| high-did | 3.46 |
| high-was | 3.32 |

Table 10: Ratings for target items in Experiment 3a (*high/low also, did-was*, with *every*)

The results show a main effect of *ellipsis size* ($p < 0.05$), such that *small ellipsis* (with *did*) was rated higher than *large ellipsis* (with *was*) across both determiners, and a main effect of *position of also* ($p < 0.05$), such that *high also* was rated higher than *low also*.

## Experiment 3b

The raw ratings for Experiment 3a are given in table 11. Recall that this experiment compares the behavior of *which* across two *ellipsis size* conditions: *small ellipsis* (*did*) and *large ellipsis* (*was*), and two positions for the potential intervener *also*: *low also*, above the embedded predicate, and *high also*, above the matrix predicate.

[35]This model did not include random slopes for the effect of *ellipsis size*, in order to allow it to converge.

| Condition | Rating |
|---|---|
| low-did | 2.75 |
| low-was | 2.55 |
| high-did | 2.88 |
| high-was | 2.76 |

Table 11: Ratings for target items in Experiment 3b (*high/low also*, *did-was*, with *which*)

The results show a main effect of *ellipsis size* ($p < 0.05$), such that *small ellipsis* (with *did*) was rated higher than *large ellipsis* (with *was*) across both determiners, and a main effect of *position of also* ($p < 0.05$), such that *high also* was rated higher than *low also*.[36]

## Discussion

We find that the strong effect of ellipsis size observed in all five self-paced reading studies is preserved in four of the five experiments shown here. In Experiment 1 we find instead an interaction, such that the *every-did* condition is rated higher than all other conditions. In addition, in Experiments 3a-b, we find a main effect of the position of *also* on ratings, such that *high also* is rated higher than *low also*. This may reflect prescriptive rules that dictate that one should not 'break an infinitive,' as the *low also* condition breaks this rule.[37]

Since there is no comprehensive study linking self-paced reading results to offline grammaticality results of the same items, it is difficult to make predictions about how the two should correlate. We therefore refrain from further interpretation of our results, and instead simply present them here to provide a comprehensive characterization of our items and to contribute to any future study which may take on the task to relate online and offline data. We simply note that nothing in these results seems to undermine the results of the self-paced reading studies reported in this paper, nor our interpretation of these results.

[36]This model did not include random slopes for the effect of *ellipsis size*, in order to allow it to converge.

[37]However, we should note that only a small number of participants in our pilot studies indicated that this affected their grammaticality judgment of our test sentences.