Locality Domains in Syntax: Evidence from Sentence Processing

Stefan Keine
University of Southern California

One of the main discoveries of generative syntax is that long-distance extraction proceeds in a successive-cyclic manner, in that these dependencies are comprised of a sequence of local extraction steps. This paper provides support for this general picture by presenting parsing evidence for the intermediate traces created by successive-cyclic movement, building on prior work by Gibson & Warren (2004). It furthermore uses this parsing evidence to investigate the distribution of intermediate traces. The central findings of this paper are that (i) there is evidence that successive-cyclic movement targets the edge of CPs, and (ii) that there is no comparable evidence for an intermediate landing site at *v*P edges. These findings are fully consistent with the classical view of successive cyclicity, according to which only finite clause edges host intermediate landing sites. In the context of phase theory, these results receive a straightforward explanation if CPs are phases, but *v*Ps are not. Two major previous arguments for *v*P phases (reconstruction in English and successive-cyclic extraction in Dinka) are reassessed and shown to not support *v*P phases after all. The processing evidence presented here provides a novel diagnostic for the distribution of phases and new evidence for their active role in online sentence processing.

1. Introduction

One of the central discoveries of the syntax of natural languages is the locality of the dependencies it creates. While displacement of an element is unbounded in principle, there is considerable evidence that, despite appearance, long movement consists of a sequence of smaller movement steps (Chomsky 1973, 1977, et seq.). Any element leaving a finite clause must first land in the highest specifer of this clause (Spec,CP in current terminology). A second movement step then moves this element from this position to its landing site in the next higher clause, as illustrated in (1). If more than one finite clause boundary is crossed, this process applies iteratively. The apparent unboundedness of displacement dependencies under this view merely reflects the unbounded number of short movement steps that such sequences may comprise.

(1) $[_{\text{CP}}$ Who$_i$ does $[_{\text{TP}}$ John think $[_{\text{CP}}$ t_i that $[_{\text{TP}}$ Mary married t_i in secret]]]]

The view that long movement proceeds successive-cyclically has been widely adopted in the generative tradition and a considerable amount of empirical evidence supporting it has been accumulated, including complementizer shift in Irish (McCloskey 1979, 2002), *wh*-copying constructions

I am particularly indebted to Brian Dillon and Lyn Frazier for critical discussions, suggestions and help throughout the various stages of this project. I would also like to thank Rajesh Bhatt, Chuck Clifton, Janet Fodor, Grant Goodall, Kyle Johnson, Ethan Poole, Shayne Sloggett, Matt Wagers, and Tessa Warren, who provided much-appreciated feedback and comments. Portions of this work were presented at CUNY 28 (University of Southern California) and WCCFL 33 (Simon Fraser University). I thank the audiences for their questions and suggestions.

(McDaniel 1989, Horvath 1997, Bruening 2006), reconstruction to intermediate trace positions (Barss 1986, Lebeaux 1988), quantifier float in West Ulster English (McCloskey 2000), and verb movement in Belfast English (Henry 1995), French (Kayne & Pollock 1978), and Iberian Spanish (Torrego 1984).

The standard minimalist explanation for the existence of successive-cyclic movement comes from the strictly cyclic nature of syntactic derivations in the form of phases. According to phase theory (Chomsky 2000, 2001), syntactic structure is periodically transferred to the interfaces and thereby rendered inaccessible for outside operations. The *Phase Impenetrability Condition* regulates that only the complement of a phase head is rendered inaccessible, whereas material in the specifier of the phase head remains accessible. Movement past a phase head must thus proceed successive-cyclically through the specifier of that phase head in order for the moving element to remain accessible.

Because the concept of a phase does not in and of itself determine which heads are phasal, the distribution of phases has been the subject of considerable interest in the recent literature. Chomsky (2000, 2001) proposes that both C and (transitive) *v* are phase heads. By the above logic, movement past either of these nodes must be successive-cyclic and both Spec,CP and Spec,*v*P must be filled with an intermediate trace (an idea that goes back to Chomsky 1986). In other words, it is not only crossclausal movement that is successive-cyclic, but also intra-clausal movement. The view that there are intraclausal landing sites in addition to Spec,CP has gained considerable popularity in recent syntactic theorizing and a number of empirical arguments have been made in its favor (see, e.g., Fox 1999, Legate 2003, Rackowski & Richards 2005, Aldridge 2008, van Urk & Richards 2015).

At the same time, this empirical support for *v*P phases is somewhat weaker than the evidence for CP phases. Chomsky's (2000, 2001) original reasons for treating *v* and C as phase heads have frequently been called into question (e.g. Bošković 2002, 2014, Epstein & Seely 2006, Boeckx & Grohmann 2007). Moreover, several of the empirical arguments in favor of a *v* phase, as well as diagnostics for phases in general, have been disputed (e.g., Matushansky 2005, den Dikken 2006, Dayal 2017).

The present paper contributes to our understanding of successive cyclicity and phases by exploring the distribution of successive cyclicity in real-time sentence processing. Building on the pioneering work of Gibson & Warren (2004), this paper presents the results of two self-paced reading experiments, which provide evidence that the effects of successive-cyclic movement—and hence phases—may be observed in online sentence processing and that such evidence may be fruitfully used to probe its distribution. This opens up a novel and largely uncharted empirical domain that can be brought to bear on long-standing questions in theoretical syntax more broadly.

Experiment 1 provides reading-time evidence for the existence of an intermediate landing site in Spec,CP, precisely what is expected if CP is a phase. Experiment 2 then compares the processing effects of CPs and *v*Ps that are crossed by movement. The results indicate a systematic difference between the two. I will argue that this difference can be immediately accounted for if CPs are targeted by intermediate movement, but *v*Ps are not. By the above logic, this suggests that CP is a phase, but *v*P is not.

The general link between the grammatical notion of successive cyclicity and evidence from language processing assumed throughout this paper is rather direct: If XP is a phase, it requires successive-cyclic movement through Spec,XP. In order for the parser to establish a grammatically licit syntactic parse, it has to postulate this intermediate trace. Evidence showing whether or not the parser postulates such a trace in the edge of a given domain thus constitutes evidence about whether this domain requires successive-cyclic movement or not.

The conclusion that CP is a phase but *v*P is not appears to contradict a number of arguments in the literature that specifically support the existence of *v*P phases (see the references above). In order to address this apparent paradox, I will discuss two particularly influential arguments in favor of *v*P phases (successive-cyclic movement in Dinka and reconstruction) and argue that both do not, after all, provide evidence for *v*P phases, a conclusion that reconciles these bodies of evidence with the conclusion reached here. These considerations raise the possibility that the mismatch between processing evidence and traditional diagnostics is only apparent.

This paper is structured as follows: Section 2 will discuss previous psycholinguistic investigations into successive cyclicity. It shows that previous evidence that has been argued to support successive cyclicity is potentially amenable to alternative explanations in terms of active gap filling. Experiment 1 in section 3 contrasts these two possible explanations and finds support for successive-cyclic movement through Spec,CP. Against this background, Experiment 2 in section 4 then investigates whether *v*Ps pattern with CPs in hosting intermediate landing sites, as predicted if both CPs and *v*Ps are phases. The results of Experiment 2 indicate that this is not the case. I then argue that this provides evidence that only CPs are phases. Section 5 reassesses two previous arguments in favor of *v*P phases. Finally, I discuss briefly the general relationship between evidence from sentence processing and that from more traditional diagnostics. Section 6 concludes.

2. Successive cyclicity in online sentence parsing

Despite the immense impact of successive cyclicity on theoretical syntax, there are very few direct investigations into the existence of successive-cyclic movement in sentence comprehension. The only designated attempts to address this question are Frazier & Clifton (1989) and Gibson & Warren (2004), both of which argue in favor of it. However, while the experimental results of these studies are compatible with successive-cyclic movement, they may also be attributed to independently motivated properties of the parser that are unrelated to successive-cyclic movement. This section will review previous results and discuss their limitations. These limitations will form the basis for Experiment 1, reported in section 3.

2.1 Previous results

The first study that explicitly argues for successive cyclicity in online sentence processing is Frazier & Clifton (1989), who argue that an unassigned filler[1] remains active across clause boundaries and that crossclausal movement dependencies are harder than monoclausal ones. Frazier & Clifton (1989) argue that these findings can be explained if (i) an intermediate landing site of the filler is responsible for carrying it across clause boundaries, and (ii) that the construal of the intermediate landing site necessary in crossclausal dependencies increases the demand such dependencies pose onto the parser.

While these interpretations are consistent with the observed results, they are not necessary. First, a filler being held active across a clause boundary does not, in and of itself, require the filler

[1] What the syntactic literature refers to as 'moved element' (or 'antecedent') and its 'trace' is usually termed the 'filler' and its 'gap', respectively, in the processing literature. These terms will be used interchangeably here.

to be linked to an intermediate trace at the edge of the lower clause. Second, movement over a clause boundary incurring more difficulty than intraclausal movement may be attributed to a variety of factors unrelated to successive cyclicity. One plausible alternative is that it is the increased distance between the filler and its gap in the two-clause condition that delays the establishment of the movement dependency and thereby increases the difficulty of parsing this structure.

A more recent investigation into processing reflexes of successive-cyclic movement is Gibson & Warren (2004) (henceforth G&W). The rationale underlying G&W's experiment is the following: There is independent evidence that the distance between the moved element and its trace is positively correlated with the difficulty with which the postulation of the gap and its semantic integration take place. In other words, the greater the distance between the moved element and its trace, the harder it is to integrate that trace (e.g., King & Just 1991, Gibson 1998, 2000, Warren & Gibson 2002, Grodner & Gibson 2005, Staub 2010, Bartek et al. 2011).[2] That the distance between the moved element and its trace should affect processing speed at the trace position is commonly assumed to follow from the fact that the semantic and morpho-syntactic features of the moved element have to be retrieved in order to successfully construe and interpret the trace. The greater the distance to this antecedent, the harder this retrieval process will be. How exactly distance is measured is subject to considerable debate in the literature, but the choice is inconsequential for the discussion here.[3]

Against this general background, G&W investigated the processing of structures like the ones in (2) using self-paced reading.

(2) a. *CP structure*

(i) *Movement*
The manager *who*$_i$ the consultant claimed **[**$_{\textbf{CP}}$ $\boxed{t_i}$ that the new proposal had <u>pleased</u> t_i **]** will hire five workers tomorrow.

(ii) *Control*
The consultant claimed that the new proposal had <u>pleased</u> the manager who will hire five workers tomorrow.

b. *DP structure*

(i) *Movement*
The manager *who*$_i$ **[**$_{\textbf{DP}}$ the consultant's claim about the new proposal **]** had <u>pleased</u> t_i will hire five workers tomorrow.

(ii) *Control*
The consultant's claim about the new proposal had <u>pleased</u> the manager who will hire five workers tomorrow.

---

[2]It is, for instance, well-known that object relative clauses are harder to process than subject relative clauses (King & Just 1991, Gordon et al. 2001, Traxler et al. 2002). One line of account attributes this contrast to the fact that the distance between the moved element and the trace is greater in object than in subject relative clauses (e.g., Just & Carpenter 1992).

[3]Metrics that have been proposed include the number of intervening words (Hawkins 1994), the number of intervening similar elements (Gordon et al. 2001, Lewis 1996), the number of intervening discourse referents (Gibson 1998, 2000, Warren & Gibson 2002, Grodner & Gibson 2005) and the time elapsed between the processing of the filler and the gap in combination with interference from similar constituent encodings (Lewis & Vasishth 2005, Vasishth & Lewis 2006). These possibilities are not mutually exclusive, of course.

In (2a.i) the relative pronoun *who* is moved across a finite CP clause boundary; in (2b.i) the same element is moved over a complex DP subject, but this movement does not cross a clause boundary. For ease of reference, I will refer to these two structures as *CP structures* and *DP structures*, respectively. The sentences in (ii) constitute the respective control structures, in which the object of the embedded verb *pleased* is not moved. Of interest are the reading times at the verb that hosts the trace of the moved element – *pleased* in (2). To obtain a measure of the difficulty of integrating the trace, G&W compared the reading times at *pleased* in the movement condition in (i) to the reading times in the corresponding non-movement structure in (ii). Because the lexical content and the immediately preceding syntactic context are identical between the two versions, an increase from the non-movement structure to the movement structure must reflect the cost of retrieving the filler and integrating the trace.

G&W reason as follows: If successive-cyclic movement proceeds through the specifier of each CP that it crosses, there will be an intermediate trace in the CP structure (boxed in (2a.i)). Because no CP is crossed in the DP structure (2b), no intermediate trace exists in this structure. Recall that the difficulty of integrating a trace depends on the distance to its antecedent. If there exists an intermediate trace in the CP condition but not in the DP condition, the distance to the antecedent should be smaller in the CP structure than in the DP structure. This in turn predicts that the integration of the trace should be easier in the CP condition than in the DP condition. On the other hand, if no intermediate trace exists in the CP condition, the integration of the trace should either be equally hard in the two conditions (if it is the linear distance between the trace and its antecedent that matters) or the integration should be harder in the CP condition than in the DP condition (if structural distance is the decisive factor). Of interest, then, is whether the integration of the trace is facilitated by a silent intermediate trace in the CP condition.[4]

G&W's results confirm the prediction of successive cyclicity. In addition to an overall reading-time increase in the movement conditions compared to the non-movement conditions, this increase was significantly smaller in the CP structure (2a) than in the DP structure (2b). G&W take this result to support the existence of an intermediate landing site in Spec,CP: The reactivation of the filler at the intermediate gap site in the CP structure aids subsequent retrieval of this filler at the ultimate gap site. This effect has subsequently been replicated by Marinis et al. (2005).[5]

2.2 Limitations

While suggestive and entirely novel, G&W's findings are not completely conclusive. As it turns out, their results are amenable to an analysis in terms of independently motivated parsing strategies that are unrelated to successive-cyclic movement. This alternative account takes the following form: A crucial feature of G&W's experimental design is that the movement dependency crosses a verb (*claim* in (2a)) in the CP structure (2a.i), but not in the DP structure (2b.i). One general concern is that the parser might have initially postulated the trace of *who* as the object of the higher verb *claim*

[4]As Chuck Clifton (p. c.) has pointed out to me, a noteworthy difference between the logic in Frazier & Clifton (1989) and Gibson & Warren (2004) is that Frazier & Clifton (1989) reason that the presence of an intermediate landing site makes processing of a structure harder, whereas Gibson & Warren (2004) reason that it makes processing at the final gap site easier.

[5]Marinis et al. (2005) also tested L2 speakers of English (L1: Chinese, Japanese, German or Greek) in addition to native speakers. Interestingly, they replicated G&W's crucial effect for L1 speakers, but not for L2 speakers. Marinis et al. (2005) argue from this result that L2 speakers underuse syntactic information in online sentence processing.

in the CP structure. Note that clear evidence that the filler originates from a lower clause becomes available only when the complementizer *that* is encountered. While thus ultimately incorrect, it is possible that the parser construes *who* as the object of the verb *claim* in its initial parse of this region. It is then furthermore possible that this intermediate, though incorrect, reactivation of *who* renders it more salient in memory and thereby facilitates its retrieval further downstream when its actual trace position is encountered. This facilitation would then be reflected as a smaller reading-time increase compared to the DP condition, where no such erroneous intermediate reactivation takes place.

On this alternative account, the effect observed by G&W would be entirely due to the particular mechanisms and decision procedures underlying the parser, and have nothing to do with successive-cyclic movement. Rather, this effect would merely be due to incorrect structural analyses temporarily pursued by the parser on the construction of a subsequent, and correct, representation. I will henceforth refer to this account as the *premature gap filling account,* as it crucially involves the premature postulation of a gap that later turns out to be incorrect.

What makes this account a viable alternative to successive cyclicity is that the pieces it comprises are independently motivated in the literature on sentence parsing. First, it is well-known that the parser pursues an *Active Filler Strategy* when scanning an input string for a gap position:

(3) *Active Filler Strategy* (Frazier & Clifton 1989: 95)
When a filler has been identified, rank the option of assigning a gap above all other options.

The active filler strategy states that the parser is extremely eager to terminate an open movement dependency by postulating a trace at the earliest grammatically licit position. Crucially, it does so even in the absence of direct evidence from the input that a trace is present. As a result, the parser may in some cases postulate a trace prematurely, i.e., in a position that will subsequently turn out to be filled by a lexical element, giving rise to 'filled gap effects', first observed by Stowe (1986). In addition to filled gap effects, a variety of other experimental paradigms have yielded support for the active filler strategy.[6]

With regard to the structures in (2), these findings are fully in line with the alternative account just sketched. Because the movement dependency crosses a higher verb (e.g, *claim*) in the CP structure, it is conceivable that the parser initially postulates the trace of *who* as the object of this higher verb. In the DP condition, on the other hand, the complex subject is an island. In light of evidence that the parser does not postulate traces in islands (Stowe 1986, Traxler & Pickering 1996, McElree & Griffith 1998), it is expected that the parser would not posit a premature trace in the complex subject in the DP structure. The ultimately incorrect but temporarily entertained trace in the CP condition would then induce a reactivation of *who* and thereby facilitate its subsequent retrieval at the actual site of the trace.

G&W briefly address this concern (p. 64), noting that the verbs they used were strongly biased towards a clausal complement and that, to the extent that these verbs were compatible with a nominal object, required their object to be inanimate. As the filler in all of their experimental items was

[6]Traxler & Pickering (1996), Phillips (2006), Staub (2007) and Wagers & Phillips (2009), for instance, find reading-time increases if a verb semantically mismatches an unassigned filler even if that filler is not syntactically related to the verb in the ultimately correct parse of the input string. There is moreover converging evidence from the stops-making-sense task (Boland et al. 1995), cross-modal lexical priming (Nicol & Swinney 1989, Nicol et al. 1994) and anticipatory eye movement (Sussman & Sedivy 2003).

animate, G&W conclude that these properties provide sufficient cues to prevent the parser from temporarily postulating a trace of the moved element at the higher verb.

Unfortunately, it is not clear at present to what extent the Active Filler Strategy can be mitigated by factors such as frequency, animacy or real-world plausibility. While subcategorization constraints have been found to affect initial parsing decisions (e.g., Staub 2007), there is evidence indicating that the frequency of a subcategorization frame does not affect whether a gap is postulated as the argument of an incoming verb (Frazier & Clifton 1989, Staub 2007) or whether a gap is postulated for an optionally transitive verb whose argument could in principle have been extraposed (Staub et al. 2006). Furthermore, Pickering & Traxler (2003) found that animacy and plausibility do not prevent the parser from construing an otherwise licit trace.

In light of these findings, it is possible that the parser postulates a trace of the moved element when it encounters the higher verb even if frequency and plausibility disfavor such a decision.[7] While G&W's results are thus fully compatible with the interpretation that the facilitation at the ultimate gap site in the CP condition is due to successive-cyclic movement, their results could potentially also be attributed to premature gap filling at the intermediate verb. To distinguish between these competing explanations, Experiment 1 investigates the role of the higher verb by systematically manipulating its subcategorization restrictions. The results favor the successive cyclicity account over the premature gap filling one.

3. Experiment 1

To distinguish whether the effect observed by Gibson & Warren (2004) is due to the integration of the filler into the intermediate Spec,CP – i.e., successive cyclicity – or to a temporarily incorrect parse in which the filler is construed as the object argument of the higher verb – i.e., premature gap filling –, Experiment 1 manipulates the subcategorization frame of the higher verb. As already mentioned, there is evidence that the parser postulates an object trace only for verbs whose subcategorization requirement allows the verb to take a DP object (Staub 2007). Under the premature gap filling account, the moved element should hence only be reactivated at the higher verb if this verb can in principle take an object DP. Consequently, the premature gap filling hypothesis predicts that the G&W effect should be modulated by properties of the intervening verb. By contrast, if G&W's effect is the result of successive cyclicity, the reactivation of the filler is independent of the higher verb and the verb type manipulation should hence not have an impact on the effect.

3.1 Method

Materials: Fifty-six sets of sentences like the one in (4) were constructed, incorporating materials adapted from Gibson & Warren (2004) and Marinis et al. (2005). Due to the fact that the semantic relations in the sentences differ by condition, a plausibility norming study was carried out to ensure that the eight conditions matched each other in plausibility. The details of this norming

[7] In addition to the general concerns just mentioned, the particular properties of G&W's stimuli also deserve some remarks. First, seven of the nine verbs used by G&W are similar to *claim* in that they productively take nominal objects (*predict, claim, conclude, imply, confirm, realize, state*). Of these seven verbs, at least four appear compatible with animate objects, as a cursory search on Google reveals: *predict* ('Jesus predicted the prophet Mohammed'), *claim* ('The queen claimed the slaves as her own property'), *confirm* ('The Senate confirmed Robert Hanna as Superior Court judge') and *realize* ('He created a trade that reached to all parts of the Union, and realized him a large fortune' [NY Times]).

study are described in Appendix A in the supplementary materials. Forty-eight sets of sentences with closely matching overall plausibility ratings were selected and the remaining eight discarded.

The experiment manipulated INTERVENER TYPE (*CP* vs. *DP*) and MOVEMENT ([+*move*] vs. [–*move*]) in a way parallel to G&W's original study. In addition, the TYPE OF THE VERB preceding the complementizer was manipulated. One set of verbs productively took nominal direct objects (like, e.g., *claim*). Verbs in this class will be referred to as *CP/DP-verbs* as their subcategorization allows for either a CP or a DP object. The second class of verbs was incompatible with a DP object and only allowed a clausal object (e.g., *think*). This second class will be referred to as *CP-verbs.* Membership in the two classes was determined by the sentence completion results reported in Trueswell et al. (1993) and a subcategorization database developed by Sabine Schulte im Walde, based on the British National Corpus and containing data for over 3000 verbs (Schulte im Walde 1998).[8] Verb type was crossed with the other two factors, yielding a total of eight conditions. Because of the relative difficulty of the stimuli, the target sentences were preceded by a theme-setting context sentence. This context sentence gave lexical information about the lower verb in the relative clause in the target sentences (e.g., *hurt* in (4)). The rationale for including this context was to decrease the processing and integration load in the critical region. Informal consultation of native speakers of American English confirmed that this context sentence reduced the perceived difficulty of the target stimuli. All target sentences were followed by a comprehension question about the semantic relations in the sentence the participant had just seen. No question targeted the context sentence. All questions were multiple choice, with both answers being referents present in the sentence. Answer options were presented in random order.

(4) *Context:* Groundless allegations really could hurt people in our company.

a. *CP structure*

(i) [+move]

The secretary *who*$_i$ the lawyer $\left\{ \begin{array}{ll} \textit{CP/DP:} & \text{claimed} \\ \textit{CP:} & \text{boasted} \end{array} \right\}$ **[**$_{\mathbf{CP}}$ that the accusations had hurt t_i **]** was fired from her job.

(ii) [–move]

The lawyer {claimed/boasted} that the accusations had hurt the secretary who was fired from her job.

[8] The 15 CP/DP-verbs used in the stimuli are *assert, assume, claim, conclude, confirm, decide, declare, demonstrate, guess, illustrate, imply, predict, prove, recall* and *state.* These verbs had a average DP object rate of .21 in the British National Corpus. The 14 CP-verbs used are *agree, argue, boast, comment, dream, hint, hope, hypothesize, insist, pretend, remark, speculate, theorize* and *think.* Of these, *agree, boast, dream, hint, hope, insist, pretend, remark* and *think* were selected because their DP object rate in Trueswell et al.'s (1993) sentence completion experiment was 0%. Five additional verbs (*argue, comment, hypothesize, speculate, theorize*) were included to decrease the amount repetition in the stimuli. Overall, these CP-verbs had a average DP object rate of .09 in the British National Corpus. It should be noted that this number is likely inflated because two verbs (*boast* and *theorize*) had a very high DP object rate (.41 and .30, respectively), which is due to the highly literary character of some of the materials in this corpus. That *boast* did not elicit a single DP object completion in Trueswell et al. (1993) makes it very unlikely that these usages of these two verbs have a significant impact on parsing decisions in the population of interest. To confirm this conclusion, all analyses reported here were also conducted with all items containing these two verb eliminated. The critical effects remained unchanged in these analyses. With these two verbs excluded, the average DP completion rate of the CP-verbs in the Corpus was .05.

b. *DP structure*

(i) [+move]

The secretary who_i **[DP** the lawyer's $\left\{\begin{array}{ll} \textit{CP/DP:} & \text{claim} \\ \textit{CP:} & \text{boast} \end{array}\right\}$ about the accusations **]** had hurt t_i was fired from her job.

(ii) [–move]

The lawyer's {claim/boast} about the accusations had hurt the secretary who was fired from her job.

Comprehension question: Who made a {claim/boast}?
the lawyer – the secretary

The difference between CP-verbs and CP/DP-verbs hence only affects the subcategorization properties of the higher verb, not the type of structure that this verb occurs in.

In addition to the 48 target items, 48 additional items were created as fillers, which matched the target sentences in syntactic complexity and length. 24 of these were part of another experiment and the additional 24 were a haphazard collection of sentences.

Participants: 130 participants were recruited via Amazon Mechanical Turk. All were native speakers of American English and naïve to the purpose of the experiment. Each received a compensation of USD 1.

Procedure: The experiment was conducted using the online experiment platform Ibex (Drummond 2013) and employed a region-by-region self-paced noncumulative moving-window task (Just et al. 1982).[9] The regioning of the target sentences followed the general schema in Table 1.[10] At the beginning of each trial participants saw the theme-setting context sentence, which was displayed in its entirety. Upon pressing the space bar, the context sentence was replaced by dashes masking the regions of the target sentences. Pressing the space bar caused the dashes in the first region to be replaced by the actual content of the region. When the space bar was pressed again this region reverted back to dashes and the next region appeared. Participants traversed through the entire sentence by repeatedly pressing the space bar. Pressing the space bar after the final region had been displayed caused the dashes to disappear and a comprehension question accompanied by two

[9]One might wonder about the reliability of reading data gathered online. Use of online platforms for this type of study is becoming increasingly mainstream among researchers (see Wagers & Phillips 2013 for a recent example). There is so far no indication that this methodology produces results that are qualitatively different from data elicited in a more traditional lab setting (e.g., Wagers & Phillips 2013). Moreover, with regard to the two experiments reported here, it is noteworthy that both extend the experimental design used by Gibson & Warren (2004) and Marinis et al. (2005) in a traditional lab setting. A measure of the reliability of the online data reported here is thus whether they replicate the key effects observed in these two studies. As emphasized in the discussion of the respective results, both Experiment 1 and 2 do replicate these previous findings. It is therefore very unlikely that the effects reported here are merely an artifact of the experimental methodology.

[10]This regioning is identical to the one employed by G&W in their analysis with one difference. In G&W's materials, relative pronouns were grouped inconsistently. In [+move] conditions the pronoun was grouped with the head noun phrase (*the secretary* in Table 1), while it was grouped with the remainder of the relative clause in the [–move] condition. This inconsistency was irrelevant for G&W as they used a word-by-word presentation and the words were grouped into regions only for the purposes of the analysis. Since, by contrast, the segmentation affects the units of presentation in the present experiment, a consistent regioning was employed and relative pronouns were presented with the head nominal throughout.

Table 1. Regioning of stimuli in Experiment 1 ([±mv] = [±move])

			Region						
			1	2	3	4	5	6	7
CP-verb	CP	[–mv]		The lawyer boasted	that	the accusations	had hurt	the secretary who	was fired from her job
		[+mv]	The secretary who	the lawyer boasted	that	the accusations	had hurt	was fired	from her job
	DP	[–mv]		The lawyer's boast	about	the accusations	had hurt	the secretary who	was fired from her job
		[+mv]	The secretary who	the lawyer's boast	about	the accusations	had hurt	was fired	from her job
CP/DP-verb	CP	[+mv]	The secretary who	the lawyer claimed	that	the accusations	had hurt	was fired	from her job
		[–mv]		The lawyer claimed	that	the accusations	had hurt	the secretary who	was fired from her job
	DP	[+mv]	The secretary who	the lawyer's claim	about	the accusations	had hurt	was fired	from her job
		[–mv]		The lawyer's claim	about	the accusations	had hurt	the secretary who	was fired from her job

possible answers appeared on the screen. Participants selected the answer on the left by pressing the 'f' key and the one on the right by pressing the 'j' key. No feedback on answer accuracy was given. After every twelve trials the participant had to take a ten-second break and could rest for longer if they so desired. Altogether, the experiment contained seven of these mandatory breaks. The experimental trials were preceded by a screen collecting general demographic data, three screens of instructions and six practice trials. The experiment took about 40 minutes. Throughout the entire experiment a progress bar was displayed. The experiment used a Latin Square design: items were arranged on eight different lists such that each list contained one instance of every item and each of the eight conditions of every item appeared on a different list. Each participants was assigned to one of the eight lists. The 48 target sentences were interspersed with the 48 fillers and the order of presentation was randomized for each participant.

Analysis: The data analysis procedures were identical for both experiments reported here and are discussed here in the context of Experiment 1.

All data analysis was carried out in the R software environment (R Core Team 2014). All reading times were logarithmically transformed and then entered into linear mixed effects (LME) models; answer accuracy was analyzed by means of logistic mixed effects modeling. Models were fit using the lme4 package (Bates et al. 2014). All models were maximal in the sense that they incorporated random intercepts for participants and items and random slopes for all fixed effects and their interactions for both participants and items, following the recommendations in Barr et al. (2013).

The estimate of the regression slope β and the corresponding t/z-statistics will be reported. To obtain p-values for the model coefficients, the Satterthwaite approximation to the degrees of freedom associated with a coefficient's t-value was employed, using the lmerTest package (Kuznetsova et al. 2014). All contrasts used in the analyses were orthogonal and will be specified for each experiment individually.

The following general exclusion procedure was applied: Participants who indicated that their native language was not American English were excluded from analysis. If a participant took the experiment multiple times, only the first time was included in the data analysis. Because of the difficulty of the materials and the subtlety of the effect of interest, only subjects whose overall accuracy over the entire experiment (including all target and filler items) was at least 75% were included in the data analysis. Reading-time data less then 200 ms or greater than 5000 ms were taken to not reflect the process of interest and hence rejected as outliers. Finally, log reading times that were more than 3 standard deviations away from the condition mean in that region were discarded.

To adjust for difference in participants' reading speed and the substantial differences in the length of the various regions, residual reading times were calculated. Using all target and filler items used in the experiment, a mixed-effects model was fit that predicted raw reading times from the number of characters in a region and included random intercepts and slopes for participants (see Ferreira & Clifton 1986 for discussion of this procedure). At every region the reading time predicted by the regression for a given participant and condition was subtracted from the actual measured reading time. This difference constitutes the residual reading time. A positive value hence indicates slower reading times than were predicted by the model while negative values predict faster reading times. Residual reading times were submitted to mixed-effects model analyses only for regions whose length differed across conditions.

3.2 Results

In order to streamline the presentation and discussion of the results, the main text will focus attention on the critical regions of the relevant experiment. A comprehensive overview of the analyses of all regions and a discussion thereof is provided in Appendix B of the supplementary materials.

Of the 130 participants who took part in Experiment 1, 6 were excluded for taking the experiment twice. An additional 30 were discarded for falling below the accuracy threshold.[11] Outlier elimination based on the absolute thresholds excluded less than 0.4% of the data. The z-score-based rejection of outliers resulted in the exclusion of 0.8% of the observations. For the analysis of the reading time data the results were subjected to a 2×2×2-factorial LME model crossing the factor MOVEMENT ([+*move*] vs. [–*move*]), INTERVENER (*CP* vs. *DP*) and VERB TYPE (*CP-verbs* vs. *CP/DP-verbs*). Accuracy data were analyzed using a parallel logistic mixed effects model. All covariates were coded numerically using sum-coding (MOVEMENT: [–*move*] = −.5, [+*move*] = .5; INTERVENER: *CP* = −.5, *DP* = .5; VERB TYPE: *CP-verb* = −.5, *CP/DP-verb* = .5). Because one item was regioned inconsistently across conditions, it was excluded from all analyses.[12]

Comprehension question response accuracy: The overall answer accuracy over all items and conditions was 85.3%. The proportion of correct responses by condition is given in Table 2. Logistic

[11]The critical statistical results reported below also hold if no accuracy-based exclusion of participants is applied.
[12]Eliminating this item did not introduce a plausibility confound (all *p*s > .5, cf. Appendix B).

Table 2. Mean answer accuracy computed over participants in Experiment 1

	CP-verbs		CP/DP-verbs	
	CP	DP	CP	DP
[–move]	.92	.93	.92	.91
[+move]	.78	.76	.83	.78

LME modeling revealed a main effect of *Movement* such that accuracy in the [+move] condition was lower than in the [–move] condition ($\hat{\beta} = -1.25, z = -11, p < .001$). No other effect was significant.

Reading times: The mean reading times by region and condition computed across participants as well as residual reading times for region which differed in length between conditions are given in Table 3. An overview of the by-condition standard error as a measure of the response variability is provided in Appendix B.

Table 4 provides an overview of the LME analyses of the log-transformed reading times for the regions that are crucial in evaluating the two contrasting hypotheses. Of central importance are the region containing the trace-hosting verb (region 5) and the spillover region (region 6). In addition, because the manipulation of the type of the higher verb should affect whether the parser prematurely posits a trace or not, the reading times in the region following this verb (region 3) should be affected by the type of the preceding verb. In the interest of space, Table 4 only reports the analyses for these three regions. A comprehensive analysis of all regions is provided in Appendix B.

Analysis of region 3 reveals main effects of *movement* and *intervener.* In addition, there was a significant three-way interaction between all factors. Because the length of the region differed between conditions, an additional analysis of the residual reading times was performed, which replicated the two main effects ($p < .001$) and the three-way interaction ($\hat{\beta} = -65, t = 2.7, p < .01$). In addition, this analysis indicated an interaction between *movement* and *intervener* ($\hat{\beta} = 25, t = 2.0, p < .05$), which is, however, uninterpretable given the higher-level interaction.

To further investigate the three-way interaction in region 3, the predictors *movement* and *verb type* were nested under the levels of *intervener* in the LME analysis of the log reading times, with the full random effects structure being preserved. The rationale for doing so is the question

Table 3. Mean raw reading times by condition and region in ms computed over participants for Experiment 1. For regions that differed in length between conditions, residual reading times are additionally provided in parentheses. (‘Vt’ = Verb type, ‘Intv’ = Intervener, ‘Mvmt’ = Movement, [±mv] = [±move])

			Region						
Vt	Intv	Mvmt	1	2	3	4	5	6	7
CP	CP	[–mv]	—	911	536 (–37)	775	708	812 (–48)	1288 (68)
		[+mv]	731	1032	598 (25)	797	836	801 (59)	868 (–64)
	DP	[–mv]	—	964	561 (–29)	824	761	796 (–62)	1213 (–10)
		[+mv]	735	1216	679 (89)	921	860	863 (120)	866 (–64)
CP/DP	CP	[–mv]	—	896	526 (–48)	765	738	840 (–19)	1279 (58)
		[+mv]	736	1066	631 (57)	790	893	820 (79)	888 (–44)
	DP	[–mv]	—	1018	562 (–20)	824	736	815 (–47)	1219 (–2)
		[+mv]	710	1215	657 (74)	892	903	845 (102)	887 (–43)

Table 4. Coefficient estimates and corresponding t-value for linear mixed effects model analyses of log reading times in critical regions of Experiment 1. *Mvmt:Intv* refers to the interaction between *Movement* and *Intervener, Mvmt:Vt* to the interaction between *Movement* and *Verb type* and *Intv:Vt* refers to the interaction between *Intervener* and *Verb type. Mvmt:Intv:Vt* refers to the three-way interaction of all predictors. Cells with $p < .05$ are shaded. For statistical analyses of all regions, see Appendix B.

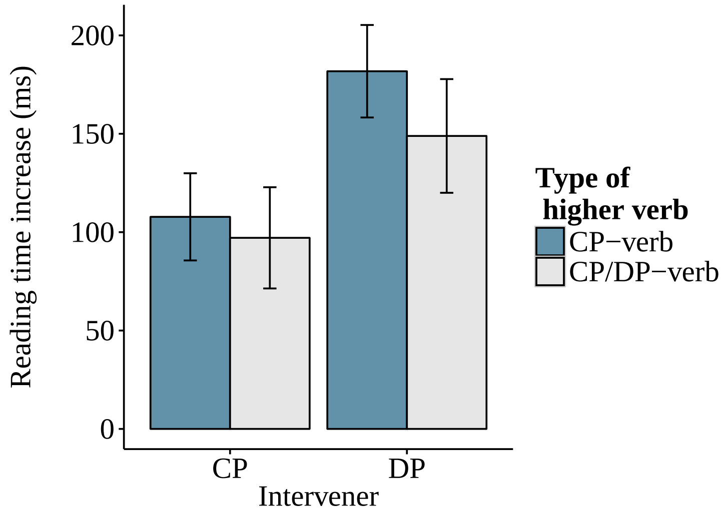
	Region					
	3		5		6	
	$\hat{\beta}$	t	$\hat{\beta}$	t	$\hat{\beta}$	t
Movement	0.127	11.33	0.112	5.03	–0.009	–0.38
Intervener	0.063	6.23	0.031	2.66	0.015	1.17
Verb type	0.000	0.03	0.023	1.99	0.012	1.11
Mvmt:Intv	0.029	1.78	–0.004	–0.17	0.060	2.55
Mvmt:Vt	0.013	0.75	0.018	0.80	–0.026	–1.27
Intv:Vt	–0.024	–1.47	–0.049	–2.37	–0.017	–0.79
Mvmt:Intv:Vt	–0.075	–2.31	0.031	0.70	–0.018	–0.43

whether CP and DP structures differ in whether or not the reading-time increase induced by movement is modulated by the type of the verb (in the CP structure) or noun (in the DP structure). For DP structures, the model detected a main effect of *movement* ($\hat{\beta} = .14, t = 10, p < .001$) but neither a main effect of *verb type* ($\hat{\beta} = -.01, t = -1.0, p = .3$) nor an interaction between the two ($\hat{\beta} = -.02, t = -1.1, p = .3$). For the CP structure, on the other hand, the model revealed a main effect of *movement* ($\hat{\beta} = .11, t = 8, p < .001$), no main effect of *verb type* ($\hat{\beta} = .01, t = 1.0, p = .3$), but an interaction between both factors ($\hat{\beta} = .05, t = 2.1, p < .05$) such that the reading-time increase was higher for CP/DP-verbs than for CP-verbs. The type of the verb/noun in region 2 thus affects the reading-times increase in region 3 in the CP structures but not the DP structures. The same pattern emerged in the analogous analyis of the residual reading times. It will be considered in greater detail in the discussion section.

In the gap region (region 5), the DP structure was read more slowly than the CP structure and the [+move] conditions were read more slowly than [–move]. There is, in addition, a main effect of *verb type* such that CP/DP-verb conditions were read more slowly than CP-verb conditions. Finally, there was an interaction between *verb type* and *intervener* such that the reading-time difference between CP and DP structures was smaller for CP/DP-verbs than for CP-verbs. There was no interaction between *movement* and *intervener*.

The spillover region (region 6) shows a significant interaction between *intervener* and *movement* such that the DP [+move] structure leads to higher reading times relative to its [–move] control than the CP structure. There is no interaction with verb type. Because the length of this region differed between the [+move] and [–move] conditions and to confirm the reliability of these critical findings, an analysis of the residual reading times was carried out. This model showed a main effect of *movement* ($\hat{\beta} = 132, t = 7, p < .001$) and an interaction between *movement* and *intervener* ($\hat{\beta} = 62, t = 2.4, p < .05$). There was no three-way interaction ($\hat{\beta} = -26, t = -.6, p > .5$). That there was a main effect of movement in only the residual reading times is unsurprising because region 6 was systematically longer in the [–move] than in the [+move] condition. Higher proportional reading

Figure 1. Mean increase in residual reading times from [–move] to [+move] conditions and by-participant standard errors in region 6 of Experiment 1 by intervener and verb type



times in the movement condition hence are only detectable when the length of the region is factored out.

Figure 1 plots the crucial increase in the residual reading times from [–move] to [+move] structures in the spillover region. The analysis just presented reveals the following key components: (i) There is a positive reading time increase incurred by the presence of a movement dependency in all conditions, (ii) this increase is greater in DP structures than in CP structures, and (iii) the pattern of this increase does not differ across the two verb types.

3.3 Discussion

The results of Experiment 1 replicate G&W's crucial finding, both in the analysis of the log-transformed and of the residual response times: The difference between [–move] and [+move] conditions was greater in DP structures than in CP structures when the position of the trace was encountered. Filler retrieval was hence more difficult in the DP structures than in the CP structure. While the region that the effect surfaced in differed (region 6 in the present experiment but region 5 in Gibson & Warren 2004 and Marinis et al. 2005), the shape of the effect is identical. Because it is common in self-paced reading experiments for an effect to surface in the spillover region, this discrepancy is not surprising.[13]

The central finding of Experiment 1 is that this effect is not modulated by the type of the higher verb, a finding that is again observed in both the log-transformed and the residualized response times. That is, the increased difficulty of retrieving the filler in the DP structure relative to the CP structure is observed regardless of the type of the higher verb. Because the type of the higher verb is crucially implicated in the premature gap filling account but not the successive cyclicity account, this result provides evidence for successive cyclicity through Spec,CP and hence G&W's interpretation of the effect. It is worth noting that the advantage of the CP condition relative to the DP condition was

[13]Note, incidentally, that the lexical material in region 6 differs in the [+move] and [–move] conditions. However, this difference in length only affects the results for the main effect of *movement*. Given that the crucial effect in this region was an interaction of the factors *movement* and *intervener* (in other words, given that we are comparing the reading-time increase between [–move] and [+move] in the CP structure to the same increase in the DP structure), which is not affected by the difference in length, this difference does not impact the interpretation of the results.

numerically, but not significantly, greater for CP-verbs than for CP/DP-verbs, the opposite of what the premature gap filling account predicts.

The three-way interaction observed in region 3 – the region containing the complementizer/preposition – is particularly instructive. The pattern of this effect is that the reading-time increase between [–move] and [+move] conditions is sensitive to the verb type in the preceding region, but only in the CP condition. More specifically, the reading-time increase is greater for CP/DP-verbs like *claim* than for CP-verbs like *boast.* In the DP condition, on the other hand, this increase is not affected by whether the preceding noun was *claim* or *boast.* This pattern is very plausibly a filled gap effect: If the verb is the CP/DP-kind, the moved element is initially construed as its object and the parser incorrectly postulates a trace of the moved element when the higher verb is encountered. The complementizer in the following region makes it clear that this parse is incorrect and that reanalysis is required. This reanalysis amplifies the reading-time increase compared to the non-movement baseline. The fact that this increase is less pronounced if the higher verb is the CP-type demonstrates that either no incorrect trace is postulated or that it is postulated on fewer trials. No such difference was observed in the DP condition: Regardless of whether the head noun is *claim* or *boast,* no gap can be postulated in either region 2 or 3. As such, no filled gap effect results because the type of the higher noun has no impact on whether or not a trace is postulated. This accounts for the observed three-way interaction in region 3. This finding has important repercussions: It demonstrates that the two groups of verbs (*CP/DP* vs. *CP*) indeed differ in their subcategorization frames and that premature gap filling is affected by these frames. Yet crucially, the critical effect in region 6 is entirely independent of these subcategorization frames. This corroborates the conclusion that the effect at the trace-hosting verb cannot be attributed to premature gap filling, as it is stable regardless of whether or not premature gap filling takes place.

There was, in addition, an interaction between *intervener* and *verb type* in the gap region such that the reading-time difference between CP and the DP structures is smaller for CP/DP-verbs than for CP-verbs. Because this pattern does not involve the reading-time increase between [–move] and [+move] structures, it is orthogonal to the crucial effects just discussed.[14]

That the presence of a movement dependency incurred an increase in reading times in regions prior to the trace is expected under the assumption that an unassigned filler has to be held active in working memory (Wanner & Maratsos 1978, Gibson 1998, 2000, Fiebach et al. 2002, Grodner et al. 2002, Chen et al. 2005), a requirement that leaves fewer resources available for the processing of incoming material.

Before proceeding, it is worth considering the role of *structural distance* in the results of Experiment 1. Taking into account structural distance does not affect the interpretation of the present results. First, suppose that no successive-cyclic movement took place in the CP condition. Because movement is crossclausal in the CP condition but intra-clausal in the DP condition, the structural

[14]It is plausible that this effect is a spillover from the preceding region and can be accounted for with reference to the definitional properties of the two verb classes: While CP/DP-verbs allow for a DP object, CP-verbs do not and it is natural to assume that this property of the verb is preserved under nominalization. In the DP structures, the noun in region 4 has to be conceptually integrated with the head noun (*claim*/*boast*). Due to subcategorization properties, this integration is likely easier for nouns based on CP/DP-verbs than for noun based on CP-verbs. In the CP condition, by contrast, the noun in region 4 is not semantically construed as an argument of the higher verb and verb type should hence not matter. As a result, the reading-time increase from the CP to the DP condition is greater for CP-verbs than for CP/DP-verbs.

distance between the filler and the gap, and hence the reading-time increase, would be predicted to be larger in the CP condition, the opposite of the observed pattern. Appeal to successive cyclicity is hence necessary even if structural distance is factored in. Moreover, if successive-cyclic movement through Spec,CP takes place, the ultimate trace is equally far away from the closest antecedent position in both conditions. Structural distance hence does not affect the conclusion that successive cyclicity takes place in the CP condition but not the DP condition and that its affect can be observed in online sentence processing.[15]

4. Experiment 2

G&W's results as well as the ones of Experiment 1 provide processing evidence for successive-cyclic movement through Spec,CP. Processing evidence thus aligns well with more traditional syntactic evidence—a welcome result. Experiment 2 uses processing evidence to ask whether intermediate traces are limited to Spec,CP (the traditional view) or whether they also occur in Spec,*v*P (the standard phase-based view). Let us refer to the former view as the *CP Hypothesis* and to the latter position as the *CP+vP Hypothesis*.

(5) a. *CP Hypothesis*
Only CP is a phase. Consequently, intermediate landing sites are created only in Spec,CP.

b. *CP+vP Hypothesis*
CP as well as *v*P are phases. Consequently, intermediate landing sites are created in Spec,CP and Spec,*v*P.

Returning to the structures investigated by G&W and in Experiment 1, the two hypotheses predict a different number of intermediate landing sites in the CP and in the DP structures. (6) and (7) give

[15] It is noteworthy that there is a second parsing-based account of the facilitation effect in the CP condition, not considered here so far. According to this account, the disproportional difficulty in the DP condition is due to a super-additive interaction between the presence of a complex subject and the existence of a movement dependency across it, the combination of which creates a computational bottleneck. Under this view, the crucial effect does not reflect particularly fast reading times in the CP condition but unexpectedly long reading times in the DP condition. In order to evaluate this account, a post hoc test was performed. Because, by hypothesis, the interaction is the super-additive combination of the relative complexity of the DP structure and the movement dependency, the size of the critical effect at region 6 should positively correlate with the complexity of the DP structure relative to the CP structure in the absence of movement. In other words, for each item, the greater the relative difficulty of the DP structure in the absence of extraction the greater the size of the interaction at region 6 is predicted to be, assuming that the complexity contributed by the movement dependency is roughly equal for all items. To evaluate this prediction, the residual reading times per structure type for regions 2 though 4 for the no-move conditions of each item were summed up. This yields a measure of the respective difficulty of the CP and DP structures in the regions preceding the critical verb in the absence of a movement dependency. To arrive at the difficulty of the DP condition relative to the CP condition, the difference of the two sums was calculated for each item. In addition, to estimate the size of the critical interaction between extraction and intervener type at region 6, the increase in log reading times from the non-extraction to the extraction condition for CP structures was subtracted from the same difference of the DP structures for each item. According to the alternative account outlined above, the size of the critical effect in region 6 should be positively correlated with the relative difficulty of the DP condition in regions 2 through 4. To assess this prediction, a linear model was devised that regressed the structural difficulty scores against the interaction scores. The model did not revealed any reliable relationship between the two measures ($r = -0.07$, *adjusted* $R^2 = -0.02$, $F(1,45) = 0.22$, $p > .5$). It is worth noting that the numerical trend was in the direction opposite of what the structural complexity account predicts. The results of Experiment 1 therefore do not support the view that the G&W effect is simply a non-cumulative combination of the two main effects.

the schematic structures and the movement steps necessary to derive them under each hypothesis. The dashed lines indicate the movement path under the CP hypothesis: there is a single intermediate gap in the CP condition and none in the DP condition. The solid lines indicate movement under the CP+*v*P hypothesis: there are three intermediate gaps in the CP structure, and one in the DP structure.

(6) *Relative clause structure in CP condition in Experiment 1*

CP hypothesis

. . . [$_{\text{CP}}$ *who* the lawyer [$_{v\text{P}}$ __ claimed [$_{\text{CP}}$ __ that the accusations had [$_{v\text{P}}$ __ hurt __]]]] . . .

CP+vP hypothesis

(7) *Relative clause structure in DP condition in Experiment 1*

CP hypothesis

. . . [$_{\text{CP}}$ *who* [$_{\text{DP}}$ the lawyer's claim about the accusations] had [$_{v\text{P}}$ __ hurt __]] . . .

CP+vP hypothesis

The results of Experiment 1 do not necessarily allow us to distinguish between the two hypotheses in (5). The results are clearly compatible with the predictions of the CP hypothesis: Because there exists an intermediate trace in the CP structure but not the DP structure, retrieval of the moved element at the final trace position is easier in the CP structure, thus giving rise to a smaller reading-time increase than in the DP structure.

Whether or not they are also compatible with the CP+*v*P hypothesis depends on the mechanics of how the moved element is reactivated over the course of the dependency. For the sake of the argument, consider the simplest version: When the ultimate trace site is encountered, the parser has to retreat to its closest antecedent – either the moved element itself, or an intermediate trace. On the CP+*v*P hypothesis, the most recent intermediate trace in both the CP and the DP structures lies in the *v*P edge immediately above the verb that hosts the trace. In other words, the distance between the ultimate trace and the closest intermediate trace is identical in the two structures in (6) and (7). If retrieving the antecedent amounted to a search for the most recent trace, there should be no difference between the two structures, contrary to fact. On this simple view of antecedent retrieval, the results of Experiment 1 favor the CP hypothesis.

There is, however, reason to believe that this picture of how the antecedent is retrieved is overly simplistic. Based on anti-locality effects in processing, Vasishth & Lewis (2006) argue that retrieval speed is also affected by the number of times an element has been previously activated (also see Lewis & Vasishth 2005). Simplifying somewhat, on their account, retrieval of an item is easier the higher the activation level of this item is. Each retrieval of an item from memory boosts that item's activation level and thereby facilitates subsequent retrievals. Applied to the structures at hand, if the representation of the moved item is retrieved from memory at each intermediate landing site, each such landing site should reactivate it. The difficulty of integrating the filler at the ultimate gap site should then be inversely related to the number of intermediate traces along the movement path. In other words, the more intermediate traces exist between the overt position of the moved item and its ultimate trace, the faster the postulation and interpretation of this trace proceeds. This view thus contrasts with the simpler one just considered in that intermediate landing sites have a *cumulative* effect on the processing speed at the trace position.[16]

Against this cumulative view of filler retrieval, let us consider again the predictions of the CP and the CP+*v*P hypothesis for the structures in (6) and (7). The predictions of the CP hypothesis remain unchanged: Because there is only a single intermediate trace in the CP structure and none in the DP structure, cumulative and non-cumulative reactivation are indistinguishable. The CP+*v*P hypothesis, by contrast, predicts that the antecedent is intermediately reactivated three times in the CP structure (corresponding to the three intermediate landing sites), but only once in the DP structure. If reactivation is cumulative, the antecedent's activation level will be greater in the CP than in the DP structure at the point at which the ultimate trace is encountered, and integration of this trace will as a result be faster in the CP structure than the DP structure. As both hypotheses thus derive the critical finding of Experiment 1, there is so far no evidence to distinguish between them if reactivation is assumed to be cumulative.

Experiment 2 is intended to differentiate between the two hypotheses by comparing the CP and DP structure to a third structure which contains an additional *v*P layer, but *no* additional CP layer. The CP hypothesis and the CP+*v*P hypothesis differ in the effect that an additional *v*P is predicted to have on retrieval times. Consider first the CP hypothesis: Because no intermediate gap is created at *v*P edges on this account, a *v*P will not lead to filler reactivation, and hence should not facilitate filler retrieval at the gap site. By contrast, the CP+*v*P hypothesis requires Spec,*v*P to contain an intermediate trace. This trace should reactivate the filler and facilitate subsequent filler retrieval at the ultimate gap site. On the CP+*v*P hypothesis, then, a *v*P layer crossed by movement should have effects entirely parallel to a CP layer. To assess these predictions, Experiment 2 investigates extraction out of *exceptional case marking* (ECM) constructions like (8), which is modeled after the sample stimuli in (6) and (7) above.[17] The central characteristic of ECM constructions is that the embedded clause is not a CP, but a TP. I will henceforth refer to structures of this type as *TP structures*. Due to the removal of the CP, the CP hypothesis predicts no intermediate landing site in (8). As two *v*Ps are crossed, the CP+*v*P hypothesis predicts two intermediate landing sites in the TP structure.

[16]In addition to the empirical evidence by Vasishth & Lewis (2006), this view is very plausible in light of current accounts of the mechanisms underlying retrieval processes in online parsing (see, among others, Stevenson 1994, McElree et al. 2003, Lewis et al. 2006, Wagers & Phillips 2009, 2013, Bartek et al. 2011). According to this family of theories, the ease with which an item is retrieved from memory is affected by this element's activation level and interference from similar encodings. An element's activation level is subject to decay over time. Retrieval of an element boosts its activation level and thereby counteracts the effects of decay and interference. As mentioned in the text, reactivation is crucially taken to be cumulative so that a sequence of reactivations boost an element's activation level to a greater extent than a single reactivation, thus aiding subsequent retrieval.

[17]ECM constructions were chosen because it is fairly uncontroversial that they lack a CP layer. Other infinitival complementation structures, like control, are commonly taken to contain a CP (e.g. Landau 2004), and are hence unsuitable to distinguish between the two hypotheses.

Note that A-movement that crosses a *v*P (as in, e.g., raising or passive constructions) is not necessarily a suitable test case because *v*Ps that do not introduce an agent argument have been argued to be non-phasal and must hence not be targeted by successive-cyclic movement (Chomsky 2000, 2001). The *v*Ps in ECM constructions, by contrast, both introduce an external argument and assign accusative case and are hence clearly phasal on Chomsky's (2000, 2001) account.

(8) *Relative clause structure in TP condition*

CP hypothesis

. . . $[_{\text{CP}}$ *who* the lawyer $[_{v\text{P}}$ __ claimed $[_{\text{TP}}$ the accusations to have $[_{v\text{P}}$ __ hurt __]]]] . . .

CP+vP hypothesis

The number of intermediate gaps postulated by the two hypotheses for each of the three structures are summarized in (9). On the null assumption that all intermediate gaps have identical effects on filler retrieval, the two hypotheses make different predictions about the difficulty of integrating the filler in the three constructions. On the CP+*v*P hypothesis, the additional *v*P layer should have the same effect as a CP layer: It should reactivate the filler and thereby facilitate its retrieval at the ultimate gap site. This hypothesis thus predicts three intermediate traces in the CP structure, two in the TP structure and one in the DP structure. The ease of integrating the trace should mirror this order: The reading-time increase should be smallest in the CP structure, larger in the TP structure and largest in the DP structure. On the CP hypothesis, on the other hand, *v*P layers should critically differ from CP layers in that only the latter facilitates filler retrieval downstream. Filler retrieval should therefore be facilitated only in the CP structure and in neither DP and TP structure, as movement does not cross a CP in the latter two.

(9) *Predicted number of intermediate traces per structure*

	CP hypothesis	CP+*v*P hypothesis
CP structure (6)	1	3
DP structure (7)	0	1
TP structure (8)	0	2
Prediction:		
Reading-time increase:	{DP, TP} > CP	DP > TP > CP

To summarize, the critical prediction of the CP hypothesis is that movement in the CP condition should be easier than in the other two, as reflected in a reading-time increase. The critical prediction of the CP+*v*P hypothesis is that the movement in the TP structure should be easier than in the DP condition.

4.1 Method

Materials: Thirty-six set of sentences like the one in (10) were constructed, out of which a subset of thirty with closely matching plausibility ratings between conditions was selected, the remaining six being discarded. The details of the norming procedure are described in Appendix A. The experiment manipulated INTERVENER TYPE (*CP* vs. *DP* vs. *TP*) and MOVEMENT ([+*move*] vs. [–*move*]). To decrease the processing load at the critical verb and to increase comparability to Experiment 1, a general scene-setting context sentence preceded each target sentence. As in Experiment 1, this context sentence gave lexical information about the embedded verbal region in the corresponding target sentence. All sentences were followed by a comprehension question about the semantic relations in the sentence. All questions were multiple choice and both possible answers occurred in the target sentence. The order of the answer options was randomized.

(10) *Context:* At the press conference last Monday several people became very agitated.

a. *CP structure*

(i) [+move]
The journalist *who*$_i$ the union member believed **[**$_{\mathbf{CP}}$ that the tax policy had intensely agitated t_i **]** was planning a series of articles.

(ii) [–move]
The union member believed that the tax policy had intensely agitated the journalist who was planning a series of articles.

b. *DP structure*

(i) [+move]
The journalist *who*$_i$ **[**$_{\mathbf{DP}}$ the union member's beliefs about the tax policy **]** had intensely agitated t_i was planning a series of articles.

(ii) [–move]
The union member's beliefs about the tax policy had intensely agitated the journalist who was planning a series of articles.

c. *TP structure*

(i) [+move]
The journalist *who*$_i$ the union member believed **[**$_{\mathbf{TP}}$ the tax policy to have intensely agitated t_i **]** was planning a series of articles

(ii) [–move]
The union member believed the tax policy to have intensely agitated the journalist who was planning a series of articles.

Comprehension question: Who believed something regarding the tax policy?
the union member – the journalist

Some general remarks about the design of the stimuli are in order. Most items contained the auxiliary *have* (*to have* in the TP condition). Other items contained the future auxiliary *will* (*to* in the TP condition). Because Experiment 2 seeks to test whether there is an intermediate landing site at the *v*P edge, it is crucial to dissociate the region containing the ultimate trace from the one containing the left edge of the *v*P. It is uncontroversial that finite auxiliaries and the infinitival *to* occupy T. An intermediate landing site in the *v*P edge can hence be reliably postulated when the parser reaches this segment of the clause. An adverb was inserted between the auxiliary and the verb hosting the trace to separate the hypothetical landing site in *v*P from the final trace position.

In addition to the 30 target sentences, the experiment contained 24 sentences of an experiment not reported here and 36 distractor sentences. All stimuli matched the target sentences in length and complexity.

Participants: 162 participants were recruited via Amazon Mechanical Turk. All were naïve to the purpose of the experiment and each received a compensation of USD 1.

Procedure: Like Experiment 1, the experiment was conducted using the online experiment platform Ibex and employed a region-by-region self-paced noncumulative moving-window task.

Table 5. Regioning of stimuli in Experiment 2 ([±mv] = [±move])

		Region								
		1	2	3	4	5	6	7	8	9
CP	[–mv]		The union member believed	that	the tax policy	had	intensely	agitated	the journalist who	was planning a series of articles
	[+mv]	The journalist who	the union member believed	that	the tax policy	had	intensely	agitated	was planning	a series of articles
DP	[–mv]		The union member's beliefs	about	the tax policy	had	intensely	agitated	the journalist who	was planning a series of articles
	[+mv]	The journalist who	the union member's beliefs	about	the tax policy	had	intensely	agitated	was planning	a series of articles
TP	[–mv]		The union member believed		the tax policy	to have	intensely	agitated	the journalist who	was planning a series of articles
	[+mv]	The journalist who	the union member believed		the tax policy	to have	intensely	agitated	was planning	a series of articles

Stimuli were regioned as shown in Table 5. The mode of presentation was identical to that in Experiment 1. Also like in Experiment 1, the participant had to take a ten-second break after every twelve trials and could rest for longer if they so desired. The total experiment thus contained seven of these breaks. The experiment took about 35 minutes. Items were arranged on six different lists such that each list contained one instance of every item and each of the six conditions of every item appeared on a different list. Each participant was assigned to one of the six lists.

Analysis: Preprocessing and outlier rejection followed the same procedure as Experiment 1.

4.2 Results

Of the 162 participants, one was excluded for not being a native speaker of English. Because of a coding error, one item and seven participants were excluded from the analysis. 53 of the remaining participants had to be excluded from analysis for falling below the accuracy threshold.[18]

Outlier rejection based on absolute reading-time thresholds eliminated less than 0.4% of the data. The z-score based outlier rejected excluded less than 1% of the data. Reading-time data were analyzed using a 2×3-factorial LME model crossing the factors MOVEMENT ([–*move*] vs. [+*move*]) and INTERVENER (*CP* vs. *DP* vs. *TP*). Accuracy data were analyzed using a parallel logistic LME model. All covariates were numerically coded. *Movement* was sum-coded ($[–move] = -.5, [+move] = .5$). The predictor *intervener* used Helmert coding: The first contrast compared the CP condition to the mean of the DP and TP conditions ($CP = -{}^2/_3, DP = {}^1/_3, TP = {}^1/_3$). The second contrast compared the DP and the TP condition ($CP = 0, TP = -.5, DP = .5$). This coding makes sense in light of the critical predictions of the two hypotheses. The first contrast tests whether the CP condition was read faster than the other two (the critical prediction of the CP hypothesis) and is reported as *IntervCP–TPDP*. The second contrast tests whether the TP condition was faster than the DP condition (the critical prediction of the CP+*v*P hypothesis) and is reported as *IntervTP–DP*. Because the TP condition had no observations in region 3, the factor *intervener* was sum-coded in this region ($CP = -.5, DP = .5$).

Comprehension question accuracy: The mean accuracy on the comprehension questions was 81%. Accuracy by condition is given in Table 6. Logistic LME modeling revealed that (i) accuracy was lower in the [+move] conditions ($\hat{\beta} = -1.6, z = -12, p < .001$); (ii) accuracy in the CP conditions was higher than the mean of the other two conditions ($\hat{\beta} = -.7, z = -4.8, p < .001$); and (iii) accuracy in the TP conditions was higher than in the DP conditions ($\hat{\beta} = -.6, z = -3.3, p < .001$). Moreover, while the accuracy drop in the [+move] conditions did not differ between CP structures and the mean of the other two ($\hat{\beta} = .3, z = 1.2, p > .2$), this drop was greater in the TP condition than in the DP condition ($\hat{\beta} = .6, z = 2.1, p < .05$). This finding will be picked up in the discussion section.

Reading times: The mean reading times per region and condition as well as the residual reading times for regions that differed in length are given in Table 7. The standard error by condition and region as a measure of data variability is provided in Appendix B.

To distinguish between the CP hypothesis and the CP+*v*P hypothesis, the crucial evidence is the

[18]As in Experiment 1, a relatively stringent accuracy criterion was chosen as the most reliable method of ensuring that the results reflect genuine properties of the parsing of the syntactic structures in questions, and not properties of shallow parsing. The critical statistical results reported here also obtain if no accuracy-based exclusion of participants is applied. The excluded item did not introduce a plausibility confound (p's $> .5$), cf. Appendix A. The coding error also led to one item being somewhat less balanced across conditions. To be conservative, analyses were conducted both with and without this item and the patterns of significance were identical in all regions. I report here the results with this item included.

pattern at the region containing the trace (region 7) and the spillover region (region 8). Of primary interest is the reading-time difference between the movement condition and the no-movement controls for each syntactic structure. Table 8 provides the results of the analyses of these two regions. In the interest of space, only the crucial gap and spillover regions are discussed in the main text. Analyses and discussions all other regions are provided in Appendix B.

In region 7 (the gap region), reading times in the TP condition were greater than in the DP condition while the CP condition did not differ from the mean of the other two. Crucially, there was an interaction between the reading-time increase due to movement noted above and the intervener type: The reading-time increase was greater in the TP condition than in the DP condition. Pairwise comparisons made it clear that this pattern was driven by an exceptionally high reading-time increase in the TP condition: The increase was greater than in the CP condition, marginally significant by participants ($t_1(100) = 1.89, p = .06$) and fully significant by items ($t_2(28) = 2.8, p < .01$) but did not differ between the CP and DP conditions ($t_1(100) = .2, p > .5;\ t_2(28) = .3, p > .5$).

In region 8 (the spillover region), the only significant effect was an interaction such that the reading-time difference between movement and non-movement structures was significantly smaller in the CP condition than in the mean of the other two. The latter two did not differ from each other. Because the length of the region differed between movement and non-movement structures, an additional LME analysis of the residual reading times was carried out to validate this effect. This analysis corroborated the interaction ($\hat{\beta} = 59, t = 2.5, p < .05$). It also showed that the [+move] conditions were read more slowly than the [–move] ones ($\hat{\beta} = 81, t = 5.2, p < .001$). Just like in Experiment 1, the fact that the region was longer in the [–move] condition than in the [+move] one renders the effect of movement visible only in the residual reading times.[19]

Figure 2 plots the crucial reading-time increase from the [–move] to the [+move] conditions in the gap and spillover regions. As the analyses just described reveal, (i) there is a positive increase in all conditions, (ii) the increase is substantially larger for TP structures in the gap region and (iii) the increase is substantially smaller for CP structures in the spillover region.

4.3 Discussion

CPs vs. *v*Ps: Recall from (9) the predictions of the two hypotheses: The critical prediction of the CP hypothesis is that a CP crossed by movement will lead to an intermediate reactivation of the

Table 6. Mean answer accuracy computed over participants by condition in Experiment 2

	CP	DP	TP
[–move]	.95	.84	.92
[+move]	.78	.67	.72

[19]Although not critical for the evaluation of the two hypotheses considered here, a post hoc analysis was carried out to compare the reading-time increase in the CP condition to that in the DP condition in the spillover region. The rationale for this test was to replicate G&W's key finding, replicated in Experiment 1 above, that the reading-time increase is greater in the DP condition than in the CP condition. Pairwise t-tests comparing the reading-time difference between [+move] and [-move] conditions in CP and DP structures were carried out on the log-transformed and residual reading times. Analysis of the former revealed that the difference was marginally smaller in the CP condition than in the DP condition ($t_1(100) = 1.62, p = .11; t_2(28) = 2.11, p < .05$). This difference was fully significant in the residual reading times ($t_1(100) = 2.23, p < .05; t_2(28) = 2.61, p < .05$). The results of Experiment 2 are thus fully consistent with those of Experiment 1.

Table 7. Mean raw reading times by condition and region in ms computed over participants for Experiment 2. For regions that differed in length between conditions, residual reading times are additionally provided in parentheses ('Intv' = Intervener, 'Mvmt' = Movement)

		Region								
Intv	Mvmt	1	2	3	4	5	6	7	8	9
CP	[–move]	—	844	494 (10)	657	493 (29)	495	522	732 (–46)	1270 (90)
	[+move]	699	949	559 (75)	717	495 (31)	502	548	677 (–2)	877 (–1)
DP	[–move]	—	1032	520 (12)	675	495 (30)	499	514	705 (–66)	1215 (38)
	[+move]	706	1190	618 (111)	770	515 (50)	507	536	714 (39)	903 (36)
TP	[–move]	—	852	—	775	543 (9)	523	525	701 (–73)	1275 (99)
	[+move]	691	1015	—	833	579 (44)	527	593	698 (26)	888 (10)

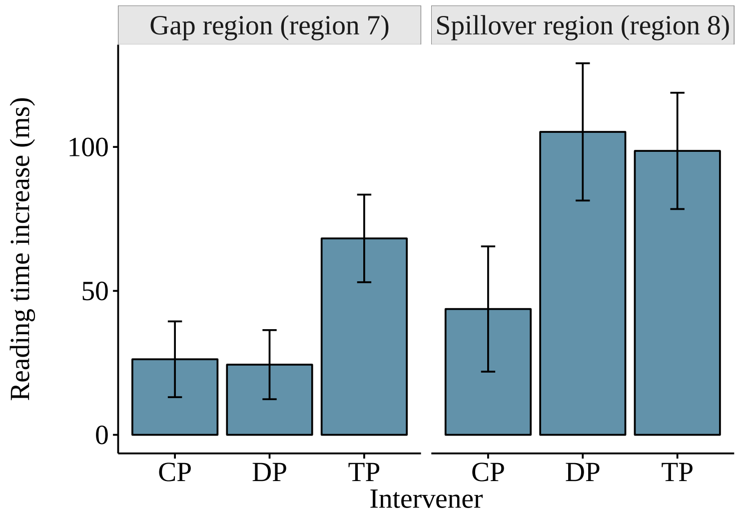
filler, while a *v*P has no such effect. On the CP+*v*P hypothesis, on the other hand, both CPs and *v*Ps are expected to reactivate the filler. The CP hypothesis thus predicts that the reading-time increase in the CP condition is smaller than in the DP and the TP condition, whereas the CP+*v*P hypothesis additionally predicts that the reading-time increase is greater in the DP condition than in the TP condition.

The results of Experiment 2 show facilitation in the CP structure relative to the other two, both in the log-transformed and of the residualized reading times. In line with the findings of Gibson & Warren (2004) and Experiment 1 above, this result provides evidence that crossing a CP layer reactivates the filler, indicating the presence of an intermediate landing site in Spec,CP. Crucially, however, there was no indication that crossing a *v*P has a similar effect. Not only was the retrieval speed in the TP condition (where two *v*Ps were crossed) not faster than in the DP control structures (where only one *v*P was crossed), retrieval times were in fact *slower* in the TP structure than in the DP structure (again both in the log-transformed and the residualized reading times). This indicates that the addition of a *v*P layer in the TP structure does *not* aid filler retrieval at the gap site, but in fact incurs a greater difficulty of filler retrieval, a result to which we will return shortly. The crucial

Table 8. Coefficient estimates and corresponding t-values for the linear mixed effects model analyses of log reading times in critical regions of Experiment 2. *Mvmt* refers to the sum-coded factor *movement; InterCP–TPDP* compares the CP condition to the mean of the TP and DP conditions and *IntervTP–DP* compares the TP condition to the DP condition. *Mvmt:IntCP–TPDP* and *Mvmt:IntTP–DP* refer to the interaction between *Mvmt* and *IntervCP–TPDP* and *IntervTP–DP*, respectively. Cells with $p < .05$ are shaded. For statistical analyses of all regions, see Appendix B.

	Region 7		Region 8	
	$\hat{\beta}$	t	$\hat{\beta}$	t
Movement	0.053	4.00	–0.039	–1.11
IntervCP–TPDP	0.013	1.13	–0.000	–0.00
IntervTP–DP	–0.042	–3.65	–0.003	–0.21
Mvmt:IntervCP–TPDP	0.024	1.09	0.059	2.08
Mvmt:IntvTP–DP	–0.049	–2.10	–0.018	–0.61

Figure 2. Mean increase in residual reading times from [–move] to [+move] conditions in gap and spillover regions and by-participant standard errors in Experiment 2 by intervener



finding of the experiment is thus that CPs and *v*Ps differ substantially with respect to the processing of a filler–gap dependency: While a CP layer leads to reactivation of the filler, a *v*P layer does not.

This differential effect of CPs and *v*Ps on filler reactivation receives a straightforward explanation if only CPs, but not *v*Ps, host intermediate landing sites, i.e., under the CP hypothesis. The observed contrast between CPs and *v*Ps is not predicted by the CP+*v*P hypothesis, on the other hand, precisely because this hypothesis treats CPs and *v*Ps on par. Thus, the hypothesis that CPs are phases and that *v*Ps are not offers a principled explanation for the results. Alternative accounts are considered below.

The results of Experiment 2 also indicate that there is no successive cyclicity in ECM constructions more generally.[20] In other words, these results not only cast doubt on successive cyclicity through Spec,*v*P, they also suggest that there is no clause-level cyclicity in TP clauses. This finding challenges recent claims that the highest projection of a clause is phasal, regardless of its category (e.g., Bošković 2014). To the extent that these conclusions are on the right track, they converge with the conclusion above that only C is a phase head (at least within the verbal domain).

The role of structural distance: While the results thus bear out the critical prediction of the CP hypothesis because only an embedded CP layer correlated with facilitated filler retrieval, the overall difficulty of the TP condition – both in the reading time and the accuracy measures – was unexpected under both hypotheses and therefore deserves discussion. Note that this increased difficulty of the TP structures relative to the DP structures directly contradicts the predictions of the CP+*v*P hypothesis, which after all predicts filler retrieval to be easier in TP structures than in DP structures. By contrast, the CP hypothesis does not by itself make a commitment with regard to the relation between the TP and DP structures. The failure of the CP hypothesis to predict the TP–DP contrast therefore does not put it on a par with the CP+*v*P hypothesis: Whereas the former merely fails to predict an observed contrast, the latter critically predicts the opposite of the observed results. Yet the question remains as to why filler retrieval should be exceptionally hard in TP structures. One plausible source of this difficulty is *structural distance.* Although the materials used in Experiment 2 controlled the *linear* distance between the overt filler and its gap, the *structural* distance between the two differed across conditions. Concretely, in the DP condition, the moved element was in the same clause as its trace (as no clause boundary was crossed in this condition), as schematized in (11) below. In the TP

[20]Thanks for Grant Goodall for pointing this out to me.

condition, on the other hand, the moved element resided in a higher clause than its trace (see (12)). Retrieval of the filler thus has to traverse a clause boundary in the TP condition. It is then possible that traversing a clause boundary for filler retrieval incurs an additional processing load. It is worth noting that this line of account converges directly with the processing study of Dillon et al. (2014), who found effects of clause boundaries in the processing of long-distance anaphora.

Importantly, this structural distance-based account of the TP–DP difference does not affect the argument for the CP hypothesis, nor is it able to reconcile the CP+*v*P hypothesis with the results. While crossing a clause boundary induces a penalty in the TP condition relative to the DP condition, it must not do so in the CP condition (or else the CP and TP conditions would pattern alike). If movement in the CP condition were crossclausal in the same way that movement in the TP condition is, they should pattern alike in being harder than the DP condition, contrary to fact. The ease of postulating the trace in the CP condition therefore shows that the structural distance is shorter in this condition than in the TP condition. This is, of course, precisely what successive cyclicity through Spec,CP claims. In other words, the penalty for crossclausal movement is absent in the CP condition precisely because there exists an intermediate trace within the same clause (see (13)). An appeal to structural distance hence does not obviate the need for successive cyclicity through Spec,CP.[21]

(11) *Movement in DP structure under CP hypothesis* ⇝ *intermediate*

[$_{\text{CP}}$ *who* [$_{\text{DP}}$ the union member's belief about the tax policy] had [$_{v\text{P}}$ intensely agitated *t*]]

(12) *Movement in TP structure under CP hypothesis* ⇝ *most difficult*

[$_{\text{CP}}$ *who* the union member [$_{v\text{P}}$ *t* believed [$_{\text{TP}}$ the tax policy to have [$_{v\text{P}}$ intensely agitated *t*]]]]

clause boundary

(13) *Movement in CP structure under CP hypothesis* ⇝ *least difficult*

[$_{\text{CP}}$ *who* the union member [$_{v\text{P}}$ believed [$_{\text{CP}}$ *t* that the tax policy had [$_{v\text{P}}$ intensely agitated *t*]]]]

clause boundary

Thus, I have suggested that the difficulty of filler retrieval in the TP structure can be given a plausible explanation if successive cyclicity through CP is combined with considerations of structural distance. It is due to the fact that (i) the filler is not intermediately reactivated (due to the lack of successive cyclicity) and that (ii) the parser has to access a higher clause to retrieve the closest antecedent of the trace. In the DP structure, (ii) does not arise. In the CP structure, neither (i) nor (ii) arises.

As mentioned above, equally importantly, a structural distance account in conjunction with the CP+*v*P hypothesis does not alter the incorrect prediction of this hypothesis. To see why, consider

[21] From the standpoint of current theories of sentence parsing, it is interesting to note that the difference between the TP and the DP conditions cannot be solely attributed to activation decay and similarity-based interference (e.g., McElree et al. 2003, Van Dyke & Lewis 2003, Lewis & Vasishth 2005, Lewis et al. 2006, Bartek et al. 2011). First, because the DP and TP structures do not differ with respect to the linear distance between the moved element and its trace, the effect of decay should be identical. Second, because the DP structure involves an additional noun (*beliefs* in (10)), similarity-based interference would predict filler retrieval to either be harder in the DP condition than in TP condition, the inverse of what was observed, or for there to be no difference at all. This indicates that there must be some other factor at play. Crossing of a clause boundary is one viable hypothesis.

again the structures in (6–8) above. On the CP+*v*P hypothesis, the intermediate trace closest to the ultimate trace site is at the edge of the embedded *v*P in all three structures. As a consequence, the structural distance between the trace and its closest antecedent is *identical* in all three structures and no clause boundary intervenes between the gap and its closest antecedent in any of the three structures. Thus, integrating structural distance into the analysis does not affect the predictions of the CP+*v*P hypothesis. In particular, it has no impact on the incorrect prediction that the integration of the trace should be easier in the TP structure than in the DP structure.

In sum, while reference to structural distance does provide an account for the difficulty in the TP condition, it does not modify the conclusion drawn above: The results of Experiment 2 are in line with the CP hypothesis because the application of structural distance is successful only on this hypothesis.[22]

Alternative interpretations: I have argued that it is the grammatical representation of the movement dependency, specifically the representation of intermediate landing sites, that underlies the reading-time results in Experiment 2. There are a number of potential alternative explanations in terms of processing principles, which warrant consideration. The first point is an alternative interpretation of the results, the latter two arise from limitations of the experimental design.

First, while the results of Experiment 2 show that only CP layers lead to facilitated retrieval of the filler at the gap site, it is a priori possible that both CPs and *v*Ps host intermediate landing sites, but that only traces in Spec,CP reactivate the filler. A difference between CP and *v*P traces along this line would reconcile the CP+*v*P hypothesis with the results of Experiment 2 as it would effectively render traces in Spec,*v*P irrelevant for the purposes of filler retrieval. While the experimental designs here do not allow us to assess this account directly, the extent to which this account constitutes a feasible alternative to the CP hypothesis depends on whether there is independent motivation for the claim that intermediate traces in CP and *v*P differ in whether or not they lead to filler reactivation. It is not evident that there is. First, such a difference could not plausibly emerge from the *syntax* of successive cyclicity because successive-cyclic movement through CP and through *v*P are formed by the same syntactic mechanism. Second, it is not clear that there is independent *processing* motivation for such a contrast. In fact, standard parsing accounts that incorporate fluctuating activation explicitly assume that retrieval of an item invariably leads to reactivation of this item (e.g., Lewis & Vasishth 2005, Vasishth & Lewis 2006) and it is not clear on this view how the filler could be retrieved at *v*P without concomitantly reactivating it.[23] I am furthermore not aware of empirical evidence for

[22]It is likewise irrelevant for the interpretation of the results whether filler reactivation is cumulative or not. As discussed on p. 17 above, cumulative reactivation refers to the hypothesis that a sequence of reactivations eases an element's subsequent retrieval to a greater extent than a single reactivation. As also discussed there, the CP+*v*P hypothesis is compatible with the results of G&W and Experiment 1 only if cumulative reactivation is adopted. However, even under cumulative reactivation the CP+*v*P hypothesis fails to account for the results of Experiment 2, as it predicts filler retrieval to be faster in the TP condition than in the DP condition (see (9)). Consequently, the CP+*v*P hypothesis makes incorrect predictions regardless of whether or not cumulative reactivation is adopted. The CP hypothesis, by contrast, makes correct predictions for both experiments reported here regardless of whether cumulative reactivation is adopted or not. This is trivially the case because on this hypothesis no structure investigated here involves more than a single intermediate landing site. Cumulative and non-cumulative reactivation are hence indistinguishable.

[23]One line of analysis could be that intermediate traces give rise to reactivation only if they reside at a clause edge, a line of account that would liken intermediate reactivation to clausal wrap-up effects. While a priori possible, it is unclear why intermediate reactivation should be principally restricted to clause boundaries. Notice, for instance, that in contrast to wrap-up effects at the end of sentences, thematic processing in the higher clause is not completed by the time the CP of the lower clause is postulated, simply because this CP is itself a thematic argument of the higher verb.

such reactivation asymmetries. In fact, clause-internal reactivation of non-moved elements is indeed attested (Vasishth & Lewis 2006) and there seems to be no reason to expect moved elements to behave differently. In sum, reconciling the results of Experiment 2 with the CP+*v*P hypothesis along these lines would require independent evidence for the existence of reactivation asymmetries of this type as well as general processing principles that predict traces in CPs and *v*Ps to differ in a way that is consistent with the results. In the absence of such independent support, I take the reading-time pattern to pose a challenge for the view that *v*Ps are phases alongside CPs.

A second point that warrants consideration is that the higher verb in Experiment 2 was a CP/DP-verb (using the terminology of Experiment 1), a direct consequence of the experimental design. Given the results of Experiment 1, it is expected that the parser initially postulates a trace of the moved element as the object of these verbs in the TP and CP conditions, but not in the DP condition. It is a priori possible that the filled-gap effect in the TP condition has affected the results at the actual gap site. The results of Experiment 2 do not allow us to directly investigate the impact of this confound, but there is reason to doubt that the crucial pattern at the gap and spillover regions are an artifact of premature gap filling at the higher verb. First, a filled gap effect would not only arise in the TP condition, but also in the CP condition. Because the reading-time increase was smallest in the CP condition, it is unclear how the filled gap effect would differentially affect only the TP condition. Second, in Experiment 1, verb type had an impact on the reading-time pattern in the region following the higher verb, but not in the trace region. This indicates that structural revision is fast and a parse temporarily pursued in the higher clause has no discernible impact on the regions of interest, in concordance with the previous literature on structural reanalysis. Third, the crucial reading-time increase is not modulated by the intervener structure in any region prior to the gap region (see Section B.2 in the supplementary materials), which equally makes it implausible that the pattern at the gap region is a sustained effect of gap filling at the higher verb. The timing of the effect thus discourages the view that the critical results in the TP condition are an artifact of premature gap filling, but a more direct exploration of this possibility would require a follow-up study.

A third noteworthy aspect of the design of Experiment 2 is that the TP structure contained a sustained ambiguity in the regions following the ECM verb. A DP following an ECM verb could be either a direct object to this verb, the subject of an ECM infinitive, or the subject of an embedded finite clause (without a complementizer). In the CP condition, on the other hand, the complementizer and the following DP disambiguate the input towards the correct structure earlier. One might wonder whether the increased difficulty in the TP structure could be attributed to this sustained ambiguity. As mentioned, the experimental design makes it impossible to directly assess this alternative. But due to the factorial design employed in the experiment, effects of baseline difficulty of the three structures are 'filtered out' in the crucial interactions. That is, the critical dependent measures are reading-time increases in a movement structure relative to a non-movement baseline of that structure. Because these structural properties of ECM constructions also arise in the non-movement baselines, these properties should not affect the reading-time increase and hence the interaction. Moreover, the time course of the processing difficulty is unexpected under this account. As just mentioned, the difficulty in the TP structure is limited to the gap and the spillover regions (regions 7 and 8). On the

---

Moreover, this account would appear to be at variance with evidence for retrieval-induced activation of a non-moved constituents occurring in the middle of a clause (e.g., Vasishth & Lewis 2006). Of course, only specific proposals can be conclusively evaluated.

ambiguity account, the ambiguous regions of the TP condition are the ones immediately following the embedding verb (region 4). The ambiguity account thus faces the challenge of explaining why the difficulty purportedly created by the structural indeterminacy does not arise until after this indeterminacy has been eliminated. By contrast, the successive cyclicity account attributes the effect to filler retrieval at the ultimate gap position and thus provides a straightforward explanation for its location.

In sum, the results of Experiment 2 show reactivation of a moved element if a CP is crossed, but no similar reactivation takes place if a *v*P is crossed. This pattern matches the critical prediction of the CP hypothesis but not that of the CP+*v*P hypothesis. The results can therefore be accounted for if landing sites are created in Spec,CP but not in Spec,*v*P. I have discussed a number of alternative interpretations of the results and demonstrated the challenges that arise for such alternatives. Investigating these alternatives directly would be a fruitful avenue for future work.

5. General discussion

5.1 The distribution of phases

The preceding sections have brought a novel type of empirical evidence to bear on the question of syntactic locality. The experimental results presented here reveal an asymmetry between CPs and *v*Ps with respect to filler retrieval: they indicate that a filler is reactivated if movement crosses a CP, but that no comparable reactivation takes place if movement crosses a *v*P. In line with the logic in Gibson & Warren (2004), I have shown that this finding can be given a principled explanation if movement creates intermediate touchdowns in CPs but not in *v*Ps, hence if CPs are phases and *v*Ps are not. If only CPs contain intermediate landing sites, it immediately follows that only the crossing of CPs has the effect of reactivating the filler and thereby speeding up processing at the gap site.

Moreover, the results suggest that there is no clause-level successive-cyclic movement in TP clauses, a finding that is likewise accounted for if only CPs are phases (in the clausal domain).

The processing evidence thus lines up closely with the substantial body of evidence from traditional diagnostics for successive-cyclic movement through Spec,CP, an encouraging result that expands the empirical evidence for successive cyclicity. On the other hand, the experiments reported here failed to detect any evidence for an analogous intermediate landing site in Spec,*v*P. This is overall consistent with the relative sparsity of direct evidence in favor of such landing sites, but it does seem to conflict with the existing evidence that uses traditional diagnostics in support of a landing sites in Spec,*v*P. It is therefore worth considering some of the arguments that have been adduced to support successive-cyclic movement through *v*P edges. While a comprehensive discussion of all arguments that have been advanced in the literature is clearly beyond the scope of this paper, it is nonetheless instructive to discuss two particularly influential ones: reconstruction in English and edge movement in Dinka. I will argue that both arguments are ultimately uncompelling in that neither entails that *v*P is a phase. This raises the interesting possibility that the discrepancy between processing and traditional evidence is only apparent.

5.2 Previous arguments for vP phases: A reassessment

5.2.1 *Reconstruction*

One classical argument comes from reconstruction and is due to Fox (1999) (also see Sauerland 2003 and Legate 2003). Fox (1999: 174) presents the sentence in (14). Here a complex DP containing a proper name (*Ms. Brown*) and a pronoun (*he*) is *wh*-moved. Crucially, this movement path crosses a pronoun coindexed with the proper name and the quantified DP *every student* binding the pronoun. This configuration imposes two constraints: First, due to Principle C, the proper name must not be c-commanded by the pronoun *her*. Second, the pronoun *he* contained inside the DP must be c-commanded by *every student* in order to be bound. Fox (1999) reasons that the DP can neither be interpreted in its surface position—because *every student* does not c-command *he* in this position—, nor in its base position—as the proper name *Ms. Brown* would be c-command by *her*, creating a Principle C violation. The only position that allows the binding requirement and Principle C to be simultaneously satisfied is situated between *every student* and *her* (indicated by a check mark in (14)). Because no CP boundary falls within this range, Fox (1999) concludes that there exists an intermediate trace within the VP/*v*P region.

(14) [Which of the books that he$_i$ asked Ms. Brown$_j$ for] did every student$_i$ _✓_ get from her$_j$ _*_ ?

Fox's (1999) argument shows that reconstruction into an intermediate position at the edge of *v*P region is *possible* and that extraction out of *v*P *can* proceed in two steps.[24] But the claim that *v*P is a phase is considerably stronger than that, in that it *forces* such movement to proceed in two steps. To demonstrate that *v*P is a phase, it would be necessary to show that extraction out of the *v*P *must* proceed through the edge of the *v*P. The example in (14) does not establish this stronger claim. Thus, (14) would also follow from the reasonable view that it is possible for one movement step to apply to the output of another. Concretely, if the *wh*-phrase can optionally undergo QR to the intermediate position, followed by *wh*-movement to its surface position, (14) is derived, irrespective of phases.[25]

It seems to me that this is a general property of evidence from reconstruction. The option of reconstructing an element into an intermediate position is uninformative with respect to the question of whether this intermediate position must be created or merely may be created. Because an argument for phases depends on establishing the former, the reconstruction facts are compatible with *v*P being a phase, but they do not provide conclusive evidence in its favor.

5.2.2 *Successive cyclicity in Dinka*

A second influential argument for *v*P phases that I will discuss here has been presented by van Urk & Richards (2015) and is based on an intriguing set of facts from the Nilotic language Dinka.[26] I will first briefly lay out the data underlying the argument for *v*P phases and then present van Urk & Richards' (2015) analysis. I will then argue that even on their own account, *v*P phases do not in fact

[24] Den Dikken (2006) argues that (14) does not even show that. Adopting Kiss' (1993) account of pair list readings of questions, wherein a universally quantified DP can covertly move above the *wh*-element, den Dikken (2006) proposes the resulting structure in (i):

(i) [every student]$_i$ [which of the books that he$_i$ asked Ms. Brown$_j$ for]$_k$ did t_i get from her$_j$ t_k

In (i), *every student* c-commands *he* and *her* is not c-commanded by *Ms. Brown*, producing the correct interpretation. No reconstruction at all is thus necessary if (i) is the LF for (14).

[25] Incidentally, Neeleman & van de Koot (2010: 346–347) develop an account on which a moving element can reconstruct into a position even if it has not moved through this position. On their account, reconstruction does not diagnose intermediate landing sites at all.

[26] I am indebted to Coppe van Urk for very helpful discussion of and help with the Dinka data.

play an analytical role. As a result, the Dinka data are uninformative as to whether *v*P is a phase or not.

Dinka has a preverbal position that must be filled by exactly one DP in regular declarative clauses. Ditransitive predicates provide the clearest illustration of this requirement. As demonstrated in the paradigm in (15), the preverbal position has to be filled by either the direct object *kìtáp* 'book' or the indirect object *Ayén* 'Ayen'. It cannot be filled by both. The other element appears to the right of the verb.

(15) a. *ɣɛ̂ɛn cé̤ ____ yiê̤n kìtáp Ayén
I PERF give book Ayen

b. *ɣɛ̂ɛn cé̤ ____ yiê̤n Ayén kìtáp
I PERF give Ayen book

c. ɣɛ̂ɛn cé̤ Ayén yiê̤n kìtáp
I PERF Ayen give book
'I gave Ayen a book.'

d. ɣɛ̂ɛn cé̤ kìtáp yiê̤n Ayén
I PERF book give Ayen
'I gave a book to Ayen.'
(van Urk & Richards 2015: 122–123)

Van Urk & Richards (2015) propose that this preverbal position is Spec,*v*P, which bears an EPP feature that requires one *v*P-internal argument to move to its specifier. Against this background, van Urk & Richards (2015) observe that movement of a *v*P-internal element to Spec,CP must proceed through this preverbal position. In (16), the indirect object *yeŋà* 'who' is *wh*-moved. In this case, the preverbal position must remain empty and cannot be filled by the direct object *kìtáp* 'book'.

(16) a. Yeŋà cí̤i môc ____ yiê̤n kìtáp?
who PERF.NS man.GEN give book
'Who did the man give the book to?'

b. *Yeŋà cí̤i môc kìtáp yiê̤n?
who PERF.NS man.GEN book give
(van Urk & Richards 2015: 125)

The same is true of *wh*-movement of the direct object (no example shown here). *Wh*-movement must thus proceed successive-cyclically through Spec,*v*P.[27] To rule out (16b), a derivation must be blocked in which *kìtáp* 'book' is moved to Spec,*v*P (as in (15d)), followed by one-fell-swoop movement of *yeŋà* 'who', as schematized in (17). Only movement via Spec,*v*P, as in (18), must be allowed.

(17) *[$_{\text{CP}}$ yeŋà$_i$ C [$_{\text{TP}}$ … [$_{v\text{P}}$ kìtáp$_j$ *v* [$_{\text{VP}}$ yiê̤n t_i t_j]]]
who book give

[27]Van Urk & Richards (2015) provide converging evidence from the stranding of a plural marker *ke* in Spec,*v*P whenever a plural element is extracted out of this *v*P. Because these stranding facts support the same empirical generalizations as those discussed in the text, I will put them aside here in the interest of space. All considerations in the text continue to hold for *ke*-stranding as well.

(18) [$_{\text{CP}}$ yeŋà$_i$ C [$_{\text{TP}}$ ... [$_{v\text{P}}$ t_i v [$_{\text{VP}}$ yiê̤n t_i kìtáb]]]]
who give book

The impossibility of the derivation in (17) appears to constitute very strong evidence that *v*P is a phase. Upon closer scrutiny, however, it turns out that this is not the case. The empirical picture is complicated by the distribution of PPs that van Urk & Richards (2015) argue are base-generated *v*P-internally, such as *yétenô* 'who'. Extraction of such PPs does not force Spec,*v*P to be empty (like (16a)), despite the fact, that *yétenô* 'who' is extracted out of the *v*P. In this case, Spec,*v*P must be filled by some element, which in (19) is *Dɛ̀ŋ* 'Deng'. Van Urk & Richards (2015) argue that movement of *yétenô* in (19) proceeds through an outer specifier of *v*Ps. The analytical challenge, then, is to ensure that such an outer specifier can only be utilized by PPs and not by DPs.

(19) Yétenô cénnè̤ Bôl Dɛ̀ŋ tɔ̀ɔc?
where PERF.OBL Bol.GEN Deng send
'Where did Bol send Deng?' (van Urk & Richards 2015: 130)

The account that van Urk & Richards (2015) propose is based on two crucial assumptions. First, they propose the language-specific featural specification of *v* in (20). According to (20), *v* contains two probes: one Case-related and one *wh*-related. The Case probe attracts a DP to the specifier of *v* (crucially not a PP), and the *wh*-probe attracts a *wh*-element. If the two probes agree with different elements, and only then, *v* has two specifiers.[28]

(20) *Feature specification of v in Dinka*
$v \begin{bmatrix} \text{Case} \\ u\text{wh} \end{bmatrix}$ (van Urk & Richards 2015: 130–131)

The second crucial assumption is the *Multitasking* principle in (21).

(21) *Multitasking* (van Urk & Richards 2015: 132)
At every step in a derivation, if a probe can trigger two operations A and B, and the features checked by A are a superset of those checked by B, the grammar prefers A.

Applied to (20), Multitasking mandates that if both the Case and the *wh*-probe can agree with the same element (i.e., if there is a *wh*-DP), then they have to. Otherwise, they agree with separate elements. If the clause does not contain a *wh*-element, then the *wh*-probe is irrelevant and the Case probe attracts a DP into Spec,*v*P.

The contrast between (17) and (18) then follows. In a configuration in which the VP contains two DP arguments, one of which is *wh*, Multitasking requires that both the Case and the *wh*-probe on *v* agree with this *wh*-DP and attract it to *v*'s specifier. With both probes checked, there is no probe that would license movement of the non-*wh* DP to a second specifier. Movement of *kìtáp* in (17) would constitute a violation of Multitasking. The *wh*-DP in Spec,*v*P then undergoes further movement to Spec,CP.

[28] Van Urk & Richards (2015) assume that the featural makeup in (20) is invariant in Dinka, but that unsuccessful probing does not crash the derivation (Preminger 2011). Thus, the *wh*-probe has to agree with a *wh*-element if there is one, but if there is not, the resulting structure is still wellformed.

In the PP-movement case in (19), no element matches both the [u*wh*] probe and the [Case] probe, and the probes can therefore agree with distinct elements without violating Multitasking. As a result, the [Case] probe attracts the non-*wh* DP to *v*'s specifier and [u*wh*] attracts the *wh*-PP to a second specifier. The *wh*-PP then undergoes further movement to Spec,CP.[29]

With this description of van Urk & Richards' (2015) system in place, let us now return to the question of whether *v*P is a phase or not. We have seen that at first glance the impossibility of the derivation in (17), repeated here in (22), in which one-fell-swoop movement out of the *v*P takes place at first glance constitutes strong evidence for *v*P phases.

(22) *[$_{\text{CP}}$ yeŋà$_i$ C [$_{\text{TP}}$ … [$_{v\text{P}}$ kìtáp$_j$ *v* [$_{\text{VP}}$ yiê̤n t_i t_j]]] [=(17)]
who book give

As it turns out, however, once Multitasking and the *v* specification in (20) are added to the account, as van Urk & Richards (2015) propose, (22) is ruled out regardless of whether *v*P is a phase or not. In (22), the *wh*-element *yeŋà* 'who' matches both [Case] and [u*wh*] on *v* and Multitasking requires both probes to agree with it as a result. *Yeŋà* 'who' therefore has to be attracted to Spec,*v*P. With both probes on *v* satisfied, *v* cannot attract another element. Moving *kìtáp* 'book' to Spec,*v*P—either instead of *yeŋà* 'who' (as in (22)) or in addition to it—constitutes a violation of Multitasking and is therefore ruled out. It follows, then, that (22) is excluded by Multitasking and the language-specific featural makeup of *v* in (20), irrespective of whether *v*P is a phase or not. In other words, the need for a extracted *wh*-element to land in Spec,*v*P on the way is already determined by the language-specific EPP property of *v* in Dinka and as such it is not necessarily diagnostic of *v*P phases. As a result, *v*P phases do not make any analytical contribution to the account. It seems, then, that van Urk & Richards' (2015) account of the Dinka facts would remain unaffected even if the assumption that *v*P is a phase were dropped. Just like in the case of the argument from reconstruction, the Dinka facts are fully compatible with *v*P phases, but they do not, after all, require them. If these considerations are on the right track, the Dinka data emerge as orthogonal to the question of whether or not *v*P is a phase.[30]

I have argued that at least some of the previous arguments in favor of *v*P phases do not, in fact, support *v*P phases. Whether or not the same can be shown for other arguments of course remains to be seen. While more work is undoubtedly required, these considerations raise the possibility that the mismatch between traditional diagnostics and the processing evidence is smaller than it initially appears.

[29] Van Urk & Richards (2015) also show that non-wh PPs cannot occupy Spec,*v*P. This follows from (20), as such PPs match neither the [*u*wh] nor the [Case] probe.

[30] As Coppe van Urk (p.c.) has pointed out to me, the assumption that *v*P is a phase might provide a meta-theoretical rationale for *why v* in Dinka has the features it does. On the other hand, however, one might invoke Occam's razor to conclude that *v*Ps are not phases, as they render the account of Dinka redundant in that with *v*P phases a whole class of derivations is simultaneously ruled out by two principles (Multitasking and phase impenetrability). Thus, it is not clear to me that either account is clearly superior on conceptual grounds.

5.3 *Successive cyclicity and a bias problem*

Before concluding this paper, I would like to point out one general limitation in traditional diagnostics for successive-cyclic movement that the use of sentence processing evidence may help overcome. As it turns out, traditional diagnostics arguably suffer from a bias problem in that they are inherently unable to detect the *absence* of an intermediate landing site. This bias problem arises because the absence of a reflex of an intermediate landing site in a position does not entail that no intermediate landing site is present in this position. Consider as an example complementizer shifts. As mentioned in section 1, in Irish, the form of a complementizer changes if movement into its specifier has taken place. In English, the form of the complementizer is not affected by extraction over it, but this of course does not entail that long extraction in English proceeds in one fell swoop. It may merely show that English lacks an overt reflex like Irish. Consequently, the absence of an overt reflex of successive cyclicity is uninformative with respect to the presence or absence of intermediate landing sites.

This point generalizes to other diagnostics for intermediate landing sites. Since Sportiche (1988), floating quantifiers have often been taken as diagnostics for intermediate landing sites (e.g., McCloskey 2000), on the view that stranding is possible in intermediate positions. As is well-known, however, *wh*-movement in English may strand a preposition in the base position, but it may not do so in an intermediate Spec,CP (Postal 1972: 213).

(23) *Who$_i$ do you believe [$_{\text{CP}}$ [$_{\text{PP}}$ to t_i]$_j$ Mary thinks [$_{\text{CP}}$ Joan talked t_j]]?

While Postal (1972) and Perlmutter & Soames (1979: 512–514) conclude from this restriction that long *wh*-movement does *not* proceed successive-cyclically, an entirely reasonable alternative response is that movement *is* successive-cyclic, but that there exists an independent constraint that bans P-stranding in Spec,CP (e.g., McCloskey 2000: 64n9). Consequently, if stranding is impossible in a given position in a given language, it again is uninformative with respect to the presence or absence of an intermediate landing site in this position.

Returning to the question of *v*P phases, I would like to contend that this inherent limitation of traditional diagnostics is significant because it entails that the claim that *v*P is a phase is virtually unfalsifiable using these diagnostics, at least within the realm of successive-cyclic movement. This is so because an argument against *v*P being a phase would have to show that extraction out of *v*P does not create an intermediate trace in Spec,*v*P. As I have just argued, traditional diagnostics are unable to to so. For example, it is not possible to strand prepositions in Spec,*v*P:

(24) *Who$_i$ did you [to t_i]$_j$ give the present t_j?

This restriction does not reveal whether there is no intermediate landing site in Spec,*v*P or whether there is but it is masked by an independent constraint on P-stranding.[31] One broader goal of the present paper was to show that considerations of processing evidence potentially allows us to overcome this bias. In Experiment 2, both the CP hypothesis and the CP+*v*P hypothesis made distinct

[31] Dayal (2017) is another good illustration of this point. She argues that Manetta's (2010) evidence for *wh*-movement to Spec,*v*P in order to remain accessible to *v*P-external material in Hindi is not compelling, but she is careful to emphasize that this refutation of Manetta's (2010) argument does not demonstrate that no such movement takes place. Another example is den Dikken's (2006) reply to Legate's (2003) arguments for *v*P phases.

and falsifiable predictions (see (9)). As such, the experiment had the ability to detect the presence *as well as* the absence of an intermediate landing site in Spec,*v*P.

Because traditional diagnostics are thus arguably unable to support a negative answer to the question of whether *v*P is a phase, the appeal of processing evidence lies in the fact that it has the potential to provide evidence both for or against intermediate landing sites in a given position. It therefore gives us an additional window into the distribution of successive cyclicity and phases, and this makes it a valuable addition to our analytical toolkit alongside more traditional diagnostics.

6. Conclusion and outlook

This paper has explored evidence from real-time sentence processing for the existence and distribution of intermediate landing sites and hence phases, building on and extending the pioneering work of Gibson & Warren (2004). First, I presented evidence that movement dependencies that cross a CP are easier to process at their gap position than dependencies that do not cross a CP. This finding is accounted for if crossing a CP requires an intermediate landing site of the moved element, which leads to reactivation of the filler, facilitating subsequent retrieval at the ultimate gap site. Experiment 1 contrasted Gibson & Warren's (2004) successive-cyclicity account with an alternative premature gap filling account, which does not resort to successive cyclicity. The results of Experiment 1 support the successive-cyclicity account. Reading-time evidence, then, provides new support for the view that CPs are phases.

Second, Experiment 2 extended this basic experimental methodology to assess whether *v*Ps have a similar facilitatory effect on the creation of the gap. The line of reasoning was that if *v*Ps host an intermediate landing site analogously to CPs, they too should induce facilitation. By contrast, if *v*Ps are not phases, they should not aid filler retrieval at the gap site and hence not induce facilitation. Experiment 2 directly contrasted these two hypotheses. The results indicate that a *v*P crossed by movement has no facilitatory effect on dependency completion, in direct contrast to CPs. This striking asymmetry between CPs and *v*Ps is of course precisely what is predicted if intermediate landing sites are created in Spec,CP, but not in Spec,*v*P, hence if CPs are phases and *v*Ps are not. Furthermore, the experiments provide evidence against clause-level cyclicity in ECM infinitives, which lack a CP projection. Overall, the experimental results provide novel support for a type of successive cyclicity that is well-established (Spec,CP), but they found no comparable evidence for a type of successive cyclicity that is less securely established independently (Spec,*v*P). Attributing phase status to CP but not *v*P allows us to account for these findings.

Third, I have reconsidered two influential arguments in support of *v*P phases and argued that closer scrutiny reveals that both do not, in fact, provide evidence for *v*P being a phase. This conclusion reconciles this evidence with the processing results presented here. As mentioned, a comprehensive reassessment of previous arguments constitutes a fruitful avenue for future work.

One speculative question that now emerges is *why* phases should be larger than they are usually taken to be. Although the results reported here do not directly inform this question, it is instructive to consider the claim in Chomsky (2000, 2001, 2005), who contends that the grammar cyclically transfers built syntactic structure in order to optimize cognitive resource demands (also see Berwick & Weinberg 1984 for a related proposal). Against this background, too many phases might in fact incur a greater processing load in the presence of a movement dependency. Because each transition from one phase to another requires an intermediate movement step, having *too many* phases might in fact incur an undue *increase* in the processing load. It is hence conceivable that a larger, C-based phase system combines the advantages of cyclic Spell-Out with a relatively small number of intermediate landing sites necessary to construct a movement dependency.

Fourth, an important virtue of processing evidence for intermediate landing sites is that it may allow us to overcome the bias problem inherent in traditional diagnostics. I have argued that traditional diagnostics of successive cyclicity are biased in that they only allow one to diagnose the presence of an intermediate landing site, not its absence. Processing evidence helps overcome this bias problem because both the claim that a given head is a phase and the claim that it is not a phase make testable and falsifiable predictions. For this reason, sentence processing provides a welcome addition to our battery of empirical diagnostics.

In spite of this virtue, I do not wish to give the impression that processing evidence is in some sense inherently superior to more traditional diagnostics. As the discussion in section 4.3 has made clear, processing evidence comes with its own limitations as it simultaneously taps into properties of the grammar and of the parser. In particular, the reasoning step from reading-time patterns to abstract properties of syntactic structures bears on inferences about both the grammar and the parser. The argument presented here is built on the assumptions that (i) the parser constructs a phrase-structure tree that obeys the wellformedness requirements of the grammar, (ii) the integration of an intermediate trace reactivates the filler in memory and thereby strengthens its representation in memory, and (iii) reactivation of an item facilitates subsequent retrieval of this item. While these three assumptions are nonexotic and well-motivated in the parsing literature, the dependence on properties of the parser is not shared by traditional evidence for successive cyclicity, but a direct consequence of using processing evidence as a window into properties of the grammar. Thus, each diagnostic—whether traditional or processing—comes with its own strengths and limitations. Theoretical hypotheses are therefore best addressed using a variety of different types of evidence, with a particular eye towards where they appear to conflict. Direct engagement with processing evidence contributes towards this goal.

References

Aldridge, Edith (2008). Phase-based account of extraction in Indonesian. *Lingua* 118: 1440–1469.

Barr, Dale, Roger Levy, Christoph Scheepers & Harry Tily (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–278.

Barss, Andrew (1986). Chains and anaphoric dependence. Ph.d. thesis, MIT, Cambridge, MA.

Bartek, Brian, Richards Lewis, Shravan Vasishth & Mason Smith (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37: 1178–1198.

Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker (2014). lme4: Linear mixed-effects models using Eigen and S4, R package version 1.1-7.

Berwick, Robert & Amy Weinberg (1984). *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition*. Cambridge, MA: MIT Press.

Boeckx, Cedric & Kleanthes Grohmann (2007). Putting phases in perspective. *Syntax* 10: 204–222.

Boland, Julie, Michael Tanenhaus, Susan Garnsey & Greg Carlson (1995). Verb argument structure in parsing and interpretation. *Journal of Memory and Language* 34: 774–806.

Bošković, Željko (2002). A-movement and the EPP. *Syntax* 5: 167–218.

Bošković, Željko (2014). Now I'm a phase, now I'm not a phase: On the variability of phases with extraction and ellipsis. *Linguistic Inquiry* 45: 27–89.

Bruening, Benjamin (2006). Differences between the wh-scope-marking and wh-copy constructions in Passamaquoddy. *Linguistic Inquiry* 37: 25–49.

Chen, Evan, Edward Gibson & Florian Wolf (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language* 52: 144–169.

Chomsky, Noam (1973). Conditions on transformations. In: *A Festschrift for Morris Halle*, ed. by Stephen Anderson & Paul Kiparsky, New York: Academic Press, pp. 232–286.

Chomsky, Noam (1977). On wh-movement. In: *Formal Syntax*, ed. by Peter Culicover, Tom Wasow & Adrian Akmajian, New York: Academic Press, pp. 71–132.

Chomsky, Noam (1986). *Barriers*. Cambridge, MA: MIT Press.

Chomsky, Noam (2000). Minimalist inquiries: The framework. In: *Step by Step: Essays in Syntax in Honor of Howard Lasnik*, ed. by Roger Martin, David Michaels & Juan Uriagereka, Cambridge, MA: MIT Press, pp. 89–155.

Chomsky, Noam (2001). Derivation by phase. In: *Ken Hale: A Life in Language*, ed. by Michael Kenstowicz, Cambridge, MA: MIT Press, pp. 1–52.

Chomsky, Noam (2005). Three factors in language design. *Linguistic Inquiry* 36: 1–22.

Dayal, Veneeta (2017). Does Hindi-Urdu have feature-driven *wh*-movement to Spec,vP? *Linguistic Inqiry* 48: 159–172.

den Dikken, Marcel (2006). A reappraisal of *v*P being phasal: A reply to Legate, Ms., CUNY.

Dillon, Brian, Wing-Yee Chow, Matthew Wagers, Taomei Guo, Fengqin Liu & Colin Phillips (2014). The structure-sensitivity of memory access: Evidence from Mandarin Chinese. *Frontiers in Psychology* 5: 1025.

Drummond, Alex (2013). Ibex Farm (version 0.3) [software]. (Retrieved from https://github.com/addrummond/ibexfarm).

Epstein, Samuel David & T. Daniel Seely (2006). *Derivations in Minimalism: Exploring the Eliminination of A-Chains and the EPP*. Cambridge: Cambridge University Press.

Ferreira, Fernanda & Charles Clifton (1986). The independence of syntactic processing. *Journal of Memory and Language* 25: 348–368.

Fiebach, Christian, Matthias Schlesewsky & Angela Friederici (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German wh-questions. *Journal of Memory and Language* 47: 250–272.

Fox, Danny (1999). Reconstruction, Binding Theory, and the interpretation of chains. *Linguistic Inquiry* 30: 157–196.

Frazier, Lyn & Charles Clifton (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4: 93–126.

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

Gibson, Edward (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In: *Image, Language, Brain*, ed. by Alec Marantz, Yasushi Miyashita & Wayne O'Neill, Cambridge, MA: MIT Press, pp. 95–126.

Gibson, Edward & Tessa Warren (2004). Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax* 7: 55–78.

Gordon, Peter, Randall Hendrick & Marcus Johnson (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27: 1411–1423.

Grodner, Daniel & Edward Gibson (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science* 29: 261–291.

Grodner, Daniel, Edward Gibson & Susanne Tunstall (2002). Syntactic complexity in ambiguity resolution. *Journal of Memory and Language* 46: 267–295.

Hawkins, John (1994). *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

Henry, Alison (1995). *Belfast English and Standard English: Dialect Variation and Parameter Setting*. Oxford: Oxford University Press.

Horvath, Julia (1997). The status of wh-expletives and the partial wh-movement construction of Hungarian. *Natural Language and Linguistic Theory* 15: 509–572.

Just, Marcel Adam & Patricia Carpenter (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99: 122–149.

Just, Marcel Adam, Patricia Carpenter & Jaqueline Woolley (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology* 111: 228–238.

Kayne, Richard & Jean Yves Pollock (1978). Stylistic inversion, successive cyclicity, and Move NP in French. *Linguistic Inquiry* 9: 595–621.

King, Jonathan & Marcel Adam Just (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language* 30: 580–602.

Kiss, Katalin É. (1993). Wh-movement and specificity. *Natural Language and Linguistic Theory* 11: 85–120.

Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen (2014). *lmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models (lmer Objects of lme4 Package).* R package version 2.0-11.

Landau, Idan (2004). The scale of finiteness and the calculus of control. *Natural Language and Linguistic Theory* 22: 811–877.

Lebeaux, David (1988). Language acquisition and the form of the grammar. Ph.D. dissertation, University of Massachusetts, Amherst, MA.

Legate, Julie Anne (2003). Some interface properties of the phase. *Linguistic Inquiry* 34: 506–516.

Lewis, Richard (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research* 25: 93–115.

Lewis, Richard & Shravan Vasishth (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29: 1–45.

Lewis, Richard, Shravan Vasishth & Julie Van Dyke (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science* 10: 447–454.

Manetta, Emily (2010). *Wh*-expletives in Hindi-Urdu: The vP phase. *Linguistic Inquiry* 41: 1–34.

Marinis, Theodore, Leah Roberts, Claudia Felser & Harald Clahsen (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition* 27: 53–78.

Matushansky, Ora (2005). Going through a phase. In: *Perspectives on Phases*, ed. by Martha McGinnis & Norvin Richards, Cambridge, MA: MITWPL, MIT Working Papers in Linguistics 49, pp. 157–181.

McCloskey, James (1979). *Transformational Syntax and Model Theoretic Semantics: A Case Study in Modern Irish*. Dordrecht: Reidel.

McCloskey, James (2000). Quantifier float and wh-movement in an Irish English. *Linguistic Inquiry* 31: 57–84.

McCloskey, James (2002). Resumptives, successive cyclicity, and the locality of operations. In: *Derivation and Explanation in the Minimalist Program*, ed. by Samuel David Epstein & T. Daniel Seely, Oxford: Blackwell, pp. 184–226.

McDaniel, Dana (1989). Partial and multiple wh-movement. *Natural Language and Linguistic Theory* 7: 565–604.

McElree, Brian, Stephanie Foraker & Lisbeth Dyer (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language* 48: 67–91.

McElree, Brian & Teresa Griffith (1998). Structural and lexical constraints on filling gaps during sentence

comprehension: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24: 432–460.
Neeleman, Ad & Hans van de Koot (2010). A local encoding of syntactic dependencies and its consequences for the theory of movement. *Syntax* 13: 331–372.
Nicol, Janet, Janet Fodor & David Swinney (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20: 1229–1238.
Nicol, Janet & David Swinney (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* 18: 5–19.
Perlmutter, David & Scott Soames (1979). *Syntactic Argumentation and the Structure of English*. Berkeley: The University of California Press.
Phillips, Colin (2006). The real-time status of island phenomena. *Language* 82: 795–823.
Pickering, Martin & Matthew Traxler (2003). Evidence against the use of subcategorization frequencies in the processing of unbounded dependencies. *Language and Cognitive Processes* 18: 469–503.
Postal, Paul (1972). Some rules that are not successive cyclic. *Linguistic Inquiry* 3: 211–222.
Preminger, Omer (2011). Agreement as a fallible operation. Ph.D. dissertation, MIT, Cambridge, MA.
R Core Team (2014). R: A language and environment for statistical computing.
Rackowski, Andrea & Norvin Richards (2005). Phase edge and extraction. *Linguistic Inquiry* 36: 565–599.
Sauerland, Uli (2003). Intermediate adjunction with A-movement. *Linguistic Inquiry* 34: 308–314.
Schulte im Walde, Sabine (1998). Automatic semantic classification of verbs according to their alternation behavior, MA thesis, Universität Stuttgart.
Sportiche, Dominique (1988). A theory of floating quantifiers and its corollaries for constituent structure. *Linguistic Inquiry* 19: 425–450.
Staub, Adrian (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33: 550–569.
Staub, Adrian (2010). Eye movement and processing difficulty in object relative clauses. *Cognition* 116: 71–86.
Staub, Adrian, Charles Clifton & Lyn Frazier (2006). Heavy NP shift is the parser's last resort: Evidence from eye movements. *Journal of Memory and Language* 54: 389–406.
Stevenson, Suzanne (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research* 23: 295–322.
Stowe, Laurie (1986). Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 1: 227–245.
Sussman, Rachel & Julie Sedivy (2003). The time course of processing syntactic dependencies. *Language and Cognitive Processes* 18: 143–163.
Torrego, Esther (1984). On inversion in Spanish and some of its effects. *Linguistic Inquiry* 15: 103–129.
Traxler, Matthew, Robin Morris & Rachel Seely (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language* 47: 69–90.
Traxler, Matthew & Martin Pickering (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35: 454–475.
Trueswell, John, Michael Tanenhaus & Christopher Kello (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19: 528–553.
Van Dyke, Julie & Richards 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 Urk, Coppe & Norvin Richards (2015). Two components of long-distance extraction: Successive cyclicity in Dinka. *Linguistic Inquiry* 46: 113–155.

Vasishth, Shravan & Richard Lewis (2006). Argument–head distance and processing complexity: Explaining both locality and antilocality effects. *Language* 82: 767–794.

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

Wagers, Matthew & Colin Phillips (2013). Going the distance: Memory and control processes in active dependency construction. *The Quarterly Journal of Experimental Psychology* : 1–31.

Wanner, Eric & Michael Maratsos (1978). An ATN approach to comprehension. In: *Linguistic Theory and Psychological Reality*, ed. by Morris Halle, Joan Bresnan & George Miller, Cambridge, MA: MIT Press, pp. 119–161.

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