

Copies and Occurrences
Bradley Larson and Norbert Hornstein

1. There are two ways of coding movement within contemporary minimalist grammar. The earliest conception, the Copy Theory of Movement (CTM), treats movement as the joint product of two more primitive operations, Copy (C) and Merge (M). Thus in a structure like (1) α “moves” from complement of β to specifier of γ by making a copy of β and merging it into the spec position of γ .

(1) [α [γ [β α]]]

Copies are tokens of a single (type) expression and group together in virtue of having the same shape (i.e. they look alike). However, given common conceptions they are unlikely to be identical. For example, it is often assumed that movement is motivated by Greed and that the features of a moving expression change (get checked/valued) through the course of a derivation.¹ If so, the identity of the token cannot be calculated with respect to every feature (as the copies will differ on some). Rather some features must count more than others or copies must group in virtue of some version of feature non-distinctness rather than identity. Details aside, what is important here is that copies are grouped in terms of some intrinsic properties; they are copies of the same thing because they share a common relevant shape.²

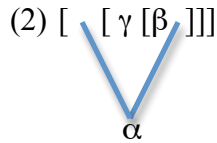
¹ See Nunes 2004 for example. This is actually more characteristic of the first generation of minimalist theory where the features that required checking were on the mover rather than the target of movement. In orthodox Probe-Goal theories, probes are unvalued and goals value them via AGREE. In such systems, movement is driven by EPP or OCC features, which have no feature checking function at all. On this conception, the copies *will* be identical as movement has no feature checking/valuation effect. We discuss the more problematic option in the text.

² There is some technology that is needed to make this run. Perhaps the most controversial is the requirement that different selections of the same lexical item from the lexicon be distinguished. Indexing is the common way of doing this. Thus if *Bob* is selected from the lexicon twice we can distinguish between *Bob1* and *Bob2*. Such indexing has been taken to violate the Inclusiveness Principle. However, if there is such indexing then it makes stating identity conditions very simple: copies are the same if they share a common index.

It is worth noting that any theory that assumes numerations must find a way of tracking multiple selections of a single lexical item. One way is to complicate the notion of a numeration so that it is not a set of lexical items but an object that is set-like but distinguishes several tokens of the same item. This is accomplished by annotating the items in the set with an index so that {John} is different from {John[2]}, the latter indicating that ‘John’ has been selected from the lexicon twice. On this conception, numerations are more complex objects than sets.

There is a second way to keep track of the number of selections and that is via indexation. So, if ‘John’ is twice selected and we index them then sets suffice to specify numerations. The analogue of {john[2]} on this implementation is {John1, John2}. Sets suffice to record the difference if selections are indexed. The main difference between these two approaches revolves around whether distinguishing the different selections is a property that persists into the syntax, as all concede it must be coded somehow in the numeration. An argument can be made that indexing simplifies the numeration while complicating the syntax, leaving the overall complexity of the grammar on either choice undetermined. Our own view is that there is not much difference between these two ways of doing things either conceptually or (at present) empirically. However, it is clear that indexing different selections from the lexicon makes it easy to determine which copies are copies of the same item and thus is the preferred method if copies are manufactured by the syntax.

The second way of coding movement bypasses the copy operation and treats it as a simple remerging of α into the spec position of γ (see Epstein et al. 1998 and Gärtner 2002 among others). We can understand this in multi-dominance-occurrence (MD-O) terms. α first merges with β and then again merges with (the projection of) γ . We could code this as in (2).



Here α has two occurrences. It is both sister of β and specifier of γ . We can identify these occurrences in terms of their sisters as Chomsky 1995 does. Thus the occurrences of α can be coded as $\langle \gamma', \beta \rangle$.³ Note, what relates occurrences here is that the same object has serially merged. What relates occurrences is their common provenance; they are all occurrences of the same expression merging again and again. There is only a single object and *it* has multiple occurrences.⁴

This paper examines the relation between copies and occurrences to see what each conceptually requires to be minimally viable given some current minimalist analyses. We will show that it is possible to smoothly translate between the Copy/ Occurrence idioms. In fact, it seems to us that these two ways of conceiving movement are rough notational variants. So, for example, one problem that the copy theory faces is which copy to pronounce. There are several interesting proposals that combine Greed (the assumption that expressions move in order to check features) with Full Interpretation (the assumption that unchecked/unvalued features crash at the interface) with bottom-up derivations to derive as consequence that *ceteris paribus* it is the highest copy that feeds the interface.⁵

³ There are other ways of doing this. For example, we can use the notions specifier and complement to do so, as we informally do in the text. We can also use immediate dominance, $\langle \gamma P, \beta P \rangle$. This gets hairy if there are multiple specifiers, if we want to distinguish outer and inner specifiers, or if we want to distinguish specifiers from adjuncts. However, we will ignore these complications in what follows and assume that there is a unique way to identify occurrences structurally.

⁴ On this conception so long as selections are kept distinct in the numeration this difference need not also persist into the syntax. What groups occurrences is *not* any kind of syntactic identity, but the fact that all are occurrences of the same expression. There is an interesting contrast made in the computational literature between deep and shallow copies that is mirrored here. With deep copies, the information from one register is duplicated in another so the information is now stored in two separate places. The copies have an independent existence and the changes in the data structures of one will not be transferred to the other. This is reminiscent of the copy operation. With shallow copies, information is kept in a single register with other registers containing pointers to the one containing the information. This is similar to what occurrences do. In this case, any change in data structure cascades to all registers that are linked via pointers. This is reminiscent of how occurrences are conceptualized. For many purposes the differences between these two conceptualizations are subtle. See, however, the discussion around (16) below.

⁵ Given other reasonable assumptions, e.g. that once all its features are checked an expression no longer moves (plausibly a consequence of Greed) and that Minimality regulates the movement of expressions with overlapping features, it is possible to derive the conclusion that *ceteris paribus* only one copy gets handed over to the interfaces for interpretation. Last, add in Recoverability, and we conclude that at least one copy must be interpreted. For a fully worked out version of this kind of theory see Nunes 2004.

An analogous (if not identical) issue faces occurrence-based accounts: which occurrence expresses the phonological content of the moved expression? It is not hard to conceive of interpretation rules that express the moved element in the position of the highest occurrence. At any rate, so far as we can tell, for the relevant empirical cases copies can be swapped for occurrences and vice versa to code the relevant data.⁶

In this short paper, we concentrate on a feature of copies that is not so readily mimicked by occurrences. The key observation is that if one copies an expression α that itself contains an expression β then when α is copied a copy of β is produced as well. This does not hold for the logic of occurrences. If α gains an occurrence, β within α need not. We will see that this is of some interest in one particular case as allowing the MD conception to work requires understanding Minimality in a very specific way, more or less as specified in Hornstein 2009. But before doing this we need to talk about how information is recycled in a derivation and how copies and occurrences are used to track this.

2 A characteristic of natural language grammars is that they are able to reuse information. Movement is a prime example. In passive sentences like (3), for example, the object *Bill* is first used as thematic object of *arrest* and is then used as the agreeing subject of the finite T.

(3) [Bill [was [arrested Bill]]]

The CTM and MD are two ways of coding this reuse. However, they do so in different ways. Let's consider some examples.

There is overwhelming empirical evidence that long distance dependencies are mediated by a series of linked successive short movements. For example, in cases of long WH movement, the relevant WH moves via a series of intermediate positions on its way to the matrix. The common analysis since the mid 1970s has been to break the long movement down into successive intermediate movements (e.g. via CP) so that the structure that results looks something like (4):

(4) [WH.....[WH.....[WH.....[WH.....WH...]]]]

⁶ It is somewhat less clear to us that the pronunciation algorithms on the market in the MD-O framework are particularly principled. There is no analogue to Greed or feature checking or Full Interpretation that regulates which occurrence is generally favored. The reason for this is that unlike Copies all occurrences have the same feature structure so the trick that Nunes 2004 plays, which says “pronounce the one with no unchecked features” cannot be played. The reason that Copies can do this is that each is a term and an independent syntactic object (potentially) differing in properties from the other copies. This is why simple identity does not suffice to group copies/tokens into types. Tokens of the same expression may not be identical in every respect as some may have features checked that others do not. This is importantly no true for occurrences. The latter have no features, only the single expression with these various relations does. As such, unlike copies, all of an all occurrences have the same feature content.

No doubt there are ways of tinkering with MD-O accounts that get around this so that the principles that drive the Copy account can be retrofitted to yield similar results when applied to occurrences. But we are not familiar with any extant proposal. We leave this issue to one side in what follows.


One standard analysis of how this structure is derived channels the CTM. Lets annotate (4) as in (5) to make this clear.

(5) [WH5.....[WH4.....[WH3.....[WH2.....WH1...]]]]

The derivation in (5) operates as follows. WH1 is copied and merged to form WH2. Similarly WH2 begets WH3, WH3 begets WH4 and WH4 begets WH5. The way that information is reused is that the *last* copy inputs to the next operations. The reuse follows a last in next out data management scheme. Note, there is no *direct* relation between WH5 and WH1 on this view of things. They relate transitively via the copy operation. Copy codes a kind of identity in that all the copies share a common relevant shape.⁷ Now, on this way of treating things, what's relevant to the next step of the WH movement operation are the properties of the immediately prior WH. To move WH2 to WH3 requires that WH2 have the relevant unchecked features (e.g. it is not specifier to a +WH C head) and it is in a position accessible to the target of movement (e.g. it is on the left edge of CP). So the next step of the movement depends on the properties of the last formed WH and the structural configuration it is in. The latter feature reflects the fact that grammatical operations are structure dependent. Note, that on this view, for example, WH1 is not actually accessible to WH3. All that is relevant for merging WH_N are the properties and structural position of WH_{N-1}. No other information concerning the WH is required to plot the next step.

This is clearly a good thing, for it codes the idea of successive cyclicity. All information required for the next step is locally available. This relies on treating all copies as *terms* in the sense of being grammatically manipulable objects. Just as WH1 can be copied and merged so too WH2, WH3....WHN.⁸

How is this coded on an MD scheme? The structure looks something like (6). The numbers are there for convenience to indicate the occurrences of the WH. Here the selfsame term sits in multiple positions.

(6) [(5)..... [(4)..... [(3)..... [(2)..... (1)...]]]]

 WH

⁷ That Copy is used to code identity is a programming truism, e.g. Marcus 2001.

⁸ The definitions in Chomsky 1995 chapter 4 treat all copies as terms and as terms are the targets of grammatical rules, all copies are expected to be grammatically manipulable.

How is successive cyclicity coded here? There are two ways. First, one can code it as a condition not on the movement operation itself but on the output chain. However, as we show below, if one adopts a movement analysis of parasitic gaps (as in Nunes 2004) this will not serve. The second option is to treat the successive merging of WH as conditioned by the accessibility of an occurrence to the target, something like: WH can merge in position N just in case one of its occurrences is accessible to N. This allows occurrences to function like copies in conditioning subsequent derivational steps. In a copy theory the feature and positional requirements necessary for computing the next derivational step are contained in the “last” copy. As all copies are terms, the logic of Minimality will suffice to restrict grammatical attention to the last merged copy. In an MD-O approach the featural and positional information are distributed between the term (which codes for the relevant features) and the occurrences of the term, which code the structural relations that the term has entered into. The upshot is that the structural sensitivity derivations enjoy can be coded by making subsequent operations involving a given term sensitive to the structural properties of its occurrences. In effect, derivations “see” a term in virtue of its occurrences and as these sit where copies do on a CTM approach, the restrictions that fit the CTM can be mimicked by an occurrence theory.⁹

3. We noted above that a representational view of occurrences that codes accessibility as a condition on the output chain appears to be a viable alternative to the derivational conception. We here present two reasons for rejecting this option, one theoretical and one empirical.

The theoretical reason relies on a minimalist assumption; that all output conditions are *bare* output conditions. We interpret this to mean that only structure that feeds interpretation persists to the CI interface. On the assumption that intermediate copies/occurrences in CP are (in general) interpretively inert and so absent at CI it is impossible to code grammatical restrictions like successive cyclicity as output conditions. This leaves the derivational option as the only viable avenue for coding this kind of locality condition.¹⁰

Empirically, an argument against treating things in output terms comes from considering Nunes 2004 analysis of parasitic gap (PG) constructions in terms of Sideways Movement (SWM), as in (7):

⁹ This said, there is something a bit odd, in our view, in so dividing grammatical responsibilities. On the standard conception concretized in Chomsky 1995, chapter 4 grammars target terms (e.g. labels not being terms are not targets of operations). It is not clear what a term is on a MD-occurrence view. The natural view is that it is the lexical item that merges and remerges. However, on a derivational view the grammatical accessibility of an expression is conditioned by the accessibility of its occurrences. Why this should be so is clear if occurrences are terms (viz. copies). It is less clear if they are not, at least in our opinion.

¹⁰ The alternative is to assume that chains in their entirety are CI objects, not just the operators and variables. This is a coherent option but one, we believe, that is not generally assumed. It is certainly incompatible with certain chain uniformity assumptions commonly made in the syntactic literature which require the elimination of “traces” in intermediate CP positions (c.f. Chomsky and Lasnik 1993). It is also possible to reject the minimalist proposal that all filters are bare output conditions. But given the minimalist setting for the discussion here we put this option to one side.

(7) What did John read *t* after he reviewed *t*

The derivation of (7) involves movement of *what* from the adjunct to the complement position of *read* before the *after* clause is adjoined. The derivation is outlined in (8):

- (8)
- a. Build the after clause and move *what* to its left edge:
[after [what [he reviewed what]]]
 - b. select *read* from the numeration and combine with *what*. This leaves two unconnected sub-trees:
[read what] [after [what [he reviewed what]]]
 - c. Merge *John* as spec of vP:
[John v [read what]] [after [what [he reviewed what]]]
 - d. Merge/adjoin *after* phrase to vP:
[[John v [read what]] [after [what [he reviewed what]]]]
 - e. Raise *John* to specT and move *What* to spec C completing the derivation:
[What C [John T [John v [read what]] [after [what [he reviewed what]]]]

There are three important features of this derivation. First, *what* is moved from the *after* phrase *before it is adjoined/merged to vP*, i.e. before the after phrase becomes an adjunct. Consequently, extracting from it does not violate the CED. Second, the movement involved is a case of SWM as it takes place between two unconnected sub-trees. The *what* is copied from the *after* phrase and merged with *read*. Third, the envelope of possible targets for SWM is a function of what a phrase Merges/adjoins to. The later the adjunction the wider the possible targets for movement.

From these three features, it follows that CED effects are a product of two factors: at what point in the derivation the adjunct is merged and what the relevant movement targets. On the (conventional) assumption that adjuncts like *after* modify some projection between vP and T', a movement that targets the complement of V will not violate the CED. Conversely, movement subsequent to adjunction of the *after* phrase will violate the CED with resulting unacceptability. This explains the deviance of sentences like (9). The target of movement is C and at the point of the derivation where C is visible, *the adjunct has been merged* and so extraction from it is illicit.¹¹

(9) *What did John read *Moby Dick* after he reviewed *t*

¹¹ Two points: There is a third relevant assumption, viz. that adjunction behaves like other forms of Merge and obeys the Extension Condition. Were this not so, then adjuncts would never be islands. Second, if this is correct we expect to find variation concerning the porosity of an adjunct depending on where it hangs (presumably a function of its semantics) and what positions the movement targets, c.f. Hornstein 2001 for discussion.

We would like to highlight a key feature of this analysis. It allows structures to vary in their island properties throughout the course of the derivation. Before an adjunct has merged it is porous and extraction is licit. After it has merged it is an island. If this is correct, then the island properties of adjuncts cannot be stated as output conditions for by the end of the derivation all adjuncts have been merged and so any dependency between the adjuncts innards and anything outside it is illicit. This is not problematic on a derivational conception if ‘being an adjunct’ is a relational notion (which it clearly is). In the context of a derivational conception, something that is an adjunct at one point of the derivation need not be so at some earlier point. This is the possibility exploited above and is what argues for a derivational perspective here.

One other point: If this is correct, then what licenses movement of *what* to Spec C in (8) is the copy/occurrence in the complement position of *read*. The other copy, being inside the adjunct cannot be manipulated without violating the CED. This is not a problem in (8) as there is a copy/occurrence that is not within an adjunct, the copy/occurrence in complement position of *read*. (9) contrasts with this in not having a copy/occurrence outside the adjoined *after* phrase. This is why movement is banned. Just as in the case of successive cyclic movement discussed earlier, so long as there is a copy/occurrence accessible to the target, the derivation is permitted.

To wrap up this section: If we analyze PGs as in Nunes 2004 then we need a derivational account of CED effects. This eliminates the option of treating locality conditions as well-formedness conditions on outputs.

4. We have noted that in a derivation the copy relevant for the next step is the last one created. Let’s grace this observation with law-like status:

- (10) In a phrase marker P, if $\alpha_n \dots \alpha_1$ are copies then if α_n is input to a subsequent step in the derivation then α_n is the most recent constructed copy.

Curiously, there is evidence that if we take (10) as regulating derivations then there is an empirical case that favors copies over occurrences. In what follows, we examine this case and discuss what feature of copies allows it to make the right distinction. We then return to question the status of (10) and show that when understood at a more fundamental level, it is quite problematic and fixing it allows for an occurrence theory to work equally well.

To get started, consider the following unacceptable sentence:

- (11) *I wonder what stories about were written by John

The unacceptability of (11) is traceable to a Subject Condition violation, *what* having been extracted from the passivized DP *stories about what*. Here’s a question: how does this account fit with contemporary views of movement. Note that in earlier GB

frameworks, the structure at the point where movement of *what* becomes an issue we have a phrase marker roughly like (12):

- (12) I wonder [_{CP} C_{+WH} [_{TP} [_{DP} stories about *what*]₁ were written t₁ by John]]

At this point in the derivation *what* resides inside a subject DP and thus comes under the purview of the Subject Condition. Note, moreover, that this is the sole WH that can move. More particularly, given that traces are not copies, there is exactly one instance of *what* that meets the requirements of the higher C_{+WH}. Consider now the CTM structure of (11) in (13):

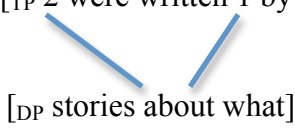
- (13) ...[_{CP} C_{+WH} [_{TP} [_{DP} stories about **what(2)**] were written [_{DP} stories about **what(1)**] by John]]

Here we have two copies of *what* and this presents a problem. Even if the Subject Condition prevents the movement of *what(2)*, what prevents moving *what(1)*?¹² Extraction from indefinite complement DPs like these is perfectly fine (as in (14)) so why can't this copy successfully feed the derivation?

- (14) What did John write stories about

The “law” in (10) suffices to block this option. If we assume that the subject DP moves before *what* moves, then according to the CTM *what(2)* is the last copy of *what* made and so it must feed the next derivational step. However, the Subject Condition prevents movement of this copy and so the derivation fails.¹³ It is worth noting how (10) helps out here. What (10) does is effectively encode the idea that all but the last copy is visible to further computational processes. It derivationally codes what trace theory coded representationally. Traces have no content. Thus they cannot feed further operations. The proposal in (10) prevents earlier copies from ever being grammatically relevant and this effectively is to treat earlier copies as if they were traces. We return to this anon.

Now consider how we could code this using an MD-occurrence approach. The structure would look something like (15) prior to WH movement:

- (15) [_{CP} C_{+WH} [_{TP} 2 were written 1 by John]]

 [_{DP} stories about *what*]

What we see here are two occurrences of *stories about what* but only one occurrence of *what*. One of these occurrences puts the DP in subject position and so the route from there to CP is plausibly illicit as it violates the Subject Condition. However,

¹² The annotating numerals are only here for ease of exposition.

¹³ The assumption that passivization precedes WH movement, though not exotic, is by no means uncontested. Chomsky 2008 argues that the movement of the two is “simultaneous.” If so, the logic outlined here will either have to be revised or abandoned.

the route from the object to CP is fine, as (14) shows. Recall that for the PG examples in (8) we assumed that licit movement required that there be at least one accessible occurrence, not that all be so.

The contrast with the CTM account in (13) revolves around there only being a single copy of *what* in (15) and thus there is no way of invoking a condition like (10). The root of this contrast lies in how copies are generated in a CTM model. There are two ways for a copy to arise. First, a copy of α can be generated if α itself is “moved.” Second a copy of α can be generated if α is contained within a larger expression β and β is moved. This is part and parcel of the CTM given the logic of copies. This logic is not endorsed in MD-O accounts. Just because β enters into two separate relations does not mean that α does. This is why (10) can be used to good effect when coupled with the CTM but not the MD-O.

This said, there is something non-aromatic about a CTM+(10) account of these instances of the subject condition, and it revolves around the status of (10). Here is our very minimalist worry: why should (10) be true? For many cases, (10) follows from more natural minimalist features. For example, if we assume that Greed (in the sense of the mover checking features at every step) drives movement, then if α is the target of movement with respect to some target, then it will be true that later copies will have more features checked than earlier ones.¹⁴ Thus, only the latest copy feed a convergent derivation. Were one to move an earlier copy then some feature would remain unchecked by the end of the derivation and it would not converge. For such cases, we get the effects of (10) without stipulating it. This is illustrated in (16) where the (un)checked features are indicated by ‘-/+'. Were we to target any but the last copy the last copy would not have all its features checked.

$$(16) \quad [\alpha^{+1/+2/+3/+4} \dots [\alpha^{+1/+2/+3/-4} \dots [\alpha^{+1/+2/-3/-4} \dots \alpha^{+1/-2/-3/-4} \dots]]]$$

For the standard cases, we can come to the same conclusion using another basic minimalist assumption. The same holds true if we assume that Extension and Minimality (or some analogues) regulate derivations. Assuming these isolates the highest copy as the favored one given that Extension guarantees c-command and Minimality will exclude all lower copies from consideration. This too is illustrated in (16). Each higher copy c-commands the lower ones and so Minimality will block the latter when further movement up the tree is called for.

The bottom line: for the standard cases of movement, when α is copied because it is the expression that is going to move, (10) is a descriptive corollary rather than a necessary axiom. This falls apart when α gets copied in virtue of being a sub-part of β . Let's consider this case now, illustrated in (17). Here α 's copy free rides on β 's having been copied. Two features of these copies of α are noteworthy. First, $\alpha(1)$ and $\alpha(2)$ have

¹⁴ We know that Greed is not a particularly important assumption anymore and that assuming that movers check features is an even less favored premise (though c.f. Boskovic 2007 for good reasons to assume it). That is why we give another route to the same conclusion below.

identical discharged and non-discharged features. Second $\alpha(1)$ does not c-command $\alpha(2)$ so Minimality is mute in such cases. In other words, for the case of free riding copies, (10) *cannot* be derived from more basic independent minimalist principles and this, we believe, is problematic, especially in the context of a derivational theory. If we assume that the input to the next step of a derivation is only the output of the last then to implement (10) we need a way of identifying which copy is “last” copy at the next step. But unless we assume that the derivational history is available (a no-no), this is not information that we have access to. This is not a problem for the standard cases of movement as the last copy coincides with the highest one and then Greed and/or Minimality will identify it as the one relevant for continuing the derivation. But when these cannot work their magic, as in (17), the grammar cannot tell simply by inspecting the last phrase marker what the right target for movement should be. In effect, $\alpha(1)$ and $\alpha(2)$ should be equally available just as they are in the MD-O accounts.

$$(17) \quad [\gamma \dots [\beta \dots \alpha(1) \dots] \dots [\beta \dots \alpha(2) \dots] \dots]$$

Now, curiously this problem with the CTM account can be fixed and brought within the purview of Minimality if we consider its inner workings a bit more.

Minimality is best seen as the injunction to create “the shortest link.”¹⁵ There is a perfectly good sense in which, if the target of α in (17) is γ that the distance between $\alpha(1)$ and γ is less than that between $\alpha(2)$ and γ . One way of codifying this is via a conception like a path, the latter being the union of the nodes dominating α and γ . In typical cases, the path from a structurally superior α will be closer to the root, viz. γ , than will be the paths from its related copies. Though, this will work, as we will presently show, some care must be taken with how the relevant nodes are counted. Illustrating this requires considering one more case.

Consider a subject parasitic gap example like (18), with the somewhat simplified structure in (19):¹⁶

- (18) (I wonder) which bill even supporters of voted for amendments to
 (19) (I wonder) [_{CP} which bill [_{TP} [_{DP1} even supporters of (which bill1)] [_{VP} voted for [_{DP2} amendments to (which bill2)]]]]]

The path from the lower copy to the higher one is {DP2, VP, TP, CP}. The path from the higher is {DP1, TP, CP}. As the latter is shorter than that former, (i.e. $|\{DP2, VP, TP, CP\}| > |\{DP1, TP, CP\}|$), movement from the real gap position should be blocked and the derivation should be ungrammatical, not a good result. Hornstein (2009) proposes that this is the wrong way to compare paths, the right way being in terms inclusion.¹⁷ So Path A is shorter than path B iff A is a proper subset of B. In the example, DP1 is distinct from DP2 and so the neither path is a proper subset of the other.

¹⁵ C.f. Chomsky and Lasnik 1993:89-90.

¹⁶ Recall, we are assuming a Nunes style SWM account of parasitic gap formation.

¹⁷ Hornstein 2009 suggests that this is plausibly tied to the fact that grammars deploy Boolean concepts, e.g. they do not count, and so measures of cardinality are unavailable.

Consequently, Minimality does not apply and movement from the real gap position is permitted.¹⁸ This is the desired result.

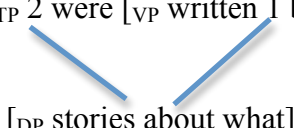
The same assumptions, however, work slightly differently in the examples like (11) discussed above. Recall that they have a structure like (13), repeated here:

(13) ...[_{CP} C+_{WH} [_{TP} [_{DP2} stories about **what(2)**] were [_{VP} written [_{DP1} stories about **what(1)**] by John]]]

The path of *what(2)* is {DP2, TP, CP}. The path of *what(1)* is {DP1, VP, TP, CP}. Note that in (13) DP2 is a copy of DP1 so the DP1 is identical to DP2. If so, however, the *what(2)* path is a proper subset of the *what(1)* path and so it is closer to C+_{WH} than *what(1)* is and so blocks access to the latter. In other words, in this particular case, where we are moving out of something that has itself moved Minimality kicks in and regulates which one the grammar can access.

In sum, if we understand Minimality in terms of paths and set inclusion we can extend (10) in a principled manner to cover (11) and (18).

However, this same apparatus can also be utilized by an MD-O approach to cover the apparently problematic (11). Consider the details. We repeat (15) here as a visual aid:

(15) [_{CP} C+_{WH} [_{TP} 2 were [_{VP} written 1 by John]]]

 [_{DP} stories about what]

Let's assume, as we did for the CTM, a notion of Minimality in terms of paths and let's assume that Minimality requires movement to occur along the shortest path. This will force *what* to move via the left branch via 2 and will block movement through the right one via 1. As the movement via 2 violates the Subject Condition the derivation is ungrammatical and unacceptability will result.

What's the upshot? There appeared to be a *prima facie* argument in favor of the CTP and against DM-O approaches when Subject Condition violations in passivized subjects were considered. However, this relied on taking (10) as a primitive condition rather than a descriptive generalization. Once we considered the theoretical reasons behind why (10) is true, we discovered that something like a minimality or shortest path requirement is implicated and this more fundamental condition seems equally applicable within a CTM or MD-O setting. In short, once again, both approaches seem equally viable and the methods of one translate straightforwardly to the other.

5. We would like to consider one last case before wrapping up.¹⁹ Chomsky (1993) considers the following case of reconstruction:

¹⁸ Minimality does not block movement from DP1 either but movement of this copy of *which bill* would result in a Subject Condition violation.

(16) John wondered which picture of himself Bill took

Chomsky 1993 observes the following correlation: when *Bill* is antecedent of the reflexive, the sentence is ambiguous between an idiomatic reading in which *take a picture* means *photograph* and one where it means *take possession of*. In contrast, when *John* is antecedent the sentence only has the second non-idiomatic interpretation.²⁰ Chomsky 1993 accounts for the contrast by considering the two structures underlying the pair of interpretations. When *John* is antecedent the structure proposed is (17):

(17) [John [₁himself₁ +T⁰] wonders [_{CP} which picture of ~~himself~~₁ [Bill took ~~which picture of himself~~]]]

The derivation supposes that in covert syntax the reflexive in CP adjoins to the higher T right next to *John*, its antecedent. At PF and LF the lower copy deletes as only one copy can surface at the LF interface. Crucially, Chomsky assumes that a necessary condition for getting the idiomatic reading is that the *picture* phrase be in the complement position at LF. This effectively relocates the venerable requirement that idioms be constituents from DS (which has been eliminated in the minimalist framework Chomsky 1993 explores) to LF (or the CI interface). Two further assumptions are required. First, that (18) is ill formed in that the distance from the lower copy of the reflexive to the matrix T⁰ is too great to be licit. Second, that (19) is not a grammatical structure, i.e. that it is not possible to retain the lower copy of *picture* while retaining the higher copy of *himself*. Chomsky 1993 does not work these assumptions out in technical detail. However, for current purposes what is relevant is the observation that which copy of the reflexive is retained depends on which copy of the head noun is interpreted at LF/CI, and, as Chomsky (1993) suggests, this can be modeled by considering the how copy deletion operates. Technical details aside, this approach seems to reify copies in the strong sense that they are actually the objects grammatically transformed. The obvious question is how to replicate this kind of analysis in an MD-O framework.

(18)[John [₁himself₁ +T⁰] wonders [_{CP} which ~~picture of himself~~ [Bill took ~~which picture of himself~~]]]

(19) [John [₁himself₁ +T⁰] wonders [_{CP} which ~~picture~~ of himself₁ [Bill took ~~which picture of himself~~]]]

The main problem for an MD-O approach is to explain the noted correlation. As far as the reflexive is concerned, one occurrence is targetable by *John* and one by *Bill*. This is not problematic. Further, one occurrence of *picture* feeds an idiomatic interpretation and one does not. Again, this is not necessarily a problem if one assumes that the contribution an expression makes to interpretation is a function of which

¹⁹ The interesting features of this example was brought to our attention by Terje Løndahl.

²⁰ It is not clear that the data as described is accurate, at least not for all speakers. At least one of the co-authors finds no difference in idiomaticity dependent on the antecedent. However, we will assume that the facts are as reported as the logic of Chomsky's proposed explanation bears on our comparison of the MTC and MD-O approaches.

occurrence is interpreted (an assumption we have made above already). What is less clear is how to link these decisions. Chomsky 1993 relies on the assumption that the occurrences have different properties and requirements. Which copy/occurrence of α is targeted determines which of β are accessible. The paper suggests that syntactic principles of deletion can exploit these to link decisions concerning them; one cannot delete the head of phrase without deleting its complement, i.e. if *picture* deletes so too must *himself* and if *himself* is retained so too must *picture* be.²¹ This makes sense if one is manipulating copies and considering how they are deleted but what is the translation of this into occurrence talk? Why should *which* occurrences get interpreted affect which other occurrences are? One can of course code this fact, but what this kind of example seems to highlight is that the grammar seems to manipulate occurrences directly (deleting some but not others) and how this is done determines what interpretations are available. Not all occurrences are on the same footing as far as the grammar is concerned. This contrasts with earlier examples. In the prior cases, occurrences conditioned operations allowing sensitivity to derivational history. However, occurrences themselves were not manipulated. Here it seems that the input to the grammatical procedures (i.e. deletion) are the occurrences themselves. This is not surprising for a CTM approach as copies being terms are manipulanda.

How to understand this if occurrences replace copies is less clear to us. Here's a suggestion. We would require a principle of the following kind:

(20) α' is an interpreted occurrence of α iff for all β' dominating α' , β' is an interpreted occurrence of β .

Something like (20) prevents “scattered” interpretations of constituents, mimicking the analogous prohibition against scattered deletions discussed above in the context of the CTM. More particularly, (20) would prohibit interpreting the higher copy of the *himself* and the lower copy of *picture* in (19). This suffices to insure the correlation between which antecedent anchored the reflexive with the availability of the idiomatic interpretation.

As noted at the outset, the option discussed here is more conceptually suggestive than empirically compelling given the caliginous nature of the relevant data. What makes this case interesting is that it might provide a testing ground for the two interpretations, though again at a very abstract level. The question will revolve around the properties of deletion processes and their relation to the kind of interpretive process mooted by (20). For the nonce, however, it seems that once again there is a way of translating (many of) the restrictions in either the CTM or MD-O idiom into the other.

As it stands, the CTM approach to the above examples captures the same data as the principle in (20): If the head is interpreted, so too must its complement. However, something like what we have in (20) cannot be the last word. Sometimes “scattered”

²¹ These in turn are plausibly related to how deletion operates. See Nunes 2004 for discussion of the costs associated with scattered deletion.

interpretations are required. An example of this can be found in what are commonly known as Lebeaux-effects (Lebeaux 1988, 1991 and Chomsky 1995). That is to say, that for the following sentences, it is still possible to get the idiomatic readings. Nevertheless, the adjunct to the wh-phrase is not interpreted low down in the idiom: There are no Principle C effects.

(21) Which picture that Bill(i) hated did he(i) say that Mary took?

(22) Which habit that Bill(i) hated did he(i) say that Mary finally kicked?

CTM accounts can rely on the late adjunction of the relative clause to capture these anti-reconstruction effects. The higher copy can be manipulated independently of the lower one and if the relative clause is appended to the higher copy only, no Principle C effects are predicted. MD-O accounts can surely also resort to late adjunction, but this is going to have an unwanted effect: the adjunct is late-adjointing low as well as high and the lack of Principle C effects is not predicted.

Perhaps the injunction against scattered interpretations in (20) could be altered so as to make reference to segments and the non-standard assumption that the relative clause adjoins to the DP 'which picture' not the NP 'picture'. In this case, the moving DP will not *fully* dominate relative clause adjunct. But even then, we could further embed the adjunct such that it would be fully dominated by something:

(23) Which picture of [the dog that Bill(i) hated] did he(i) say that Mary took?

What injunction could be made such that an MD-O account could handle these anti-reconstruction effects? Perhaps a “whatever works” sort of approach that would absolve any Binding Principle violation if at least one occurrence obeyed them. This would falsely predict that non-adjunctions would also show anti-reconstruction effects:

(24) *Which picture of Bill(i) did he(i) say that Mary took?

One could claim that idiom-internal adjuncts are immune to binding theory strictures. Though this would also make the incorrect predictions:

(25) *He(i) took some pictures that Bill(i) hated.

Perhaps for lack of imagination, we find it difficult to state a principle to capture these facts in an MD-O idiom. Adjuncts will always be dominated by an interpreted occurrence. When that interpreted occurrence must be low, there is no escape from the Principle C effects: an undesirable result. We leave these facts that couple anti-reconstruction effects together with idiom interpretations as a puzzle for MD-O theories.

5. It is a discovery of Generative Grammar that grammatical operations are structure dependent. Grammars manipulate terms subject to structural constraints. In this short paper, we have aimed to understand how two apparently distinct formal approaches to movement within MP code this fact. The CTM respects structure dependence by treating copies as terms. Like any term, a copy has a feature matrix and occupies a position in a phrase marker. Structure dependence is coded in the fact that its grammatical potency is a function both of the feature specifications the copy carries and the structural position the copy finds itself in. In MD-O accounts things are a bit more complex. These require a redefinition of the notion 'term.' Unlike within the CTM a term cannot be identified with an occurrence, as the latter is just a context (e.g. as in Chomsky 1995). Structure dependence on this conception is coded in making the *accessibility* of a term to a grammatical operation sensitive to structural features of its occurrences. In this paper we have considered a variety of cases where distinguishing copies matters and considered how to translate Copy talk into Occurrence talk without empirical loss. We have shown that this translation is by and large efficacious. There remains one case where we have been unable to make a smooth translation between the two approaches, but given its singularity we are loath to conclude that MD-O approaches are inferior to CTM analyses. In sum, the above review has led us to the (cautious) conclusion that though they appear to differ technically the CTM and MD-O idioms are largely interchangeable. This suggests that these are, at least at present, notational variants. We suspect that these two ways of conceiving the problem of the grammatical computation is akin to the different ways of understanding computation more generally. It is well known that Turing and Church converged on what appear to be a similar idea of computability, but with rather different technologies. Turing deployed tapes, copying procedures, inscription procedures, tape movements and other primitive operations to define computability. Church selected a class of primitive computations and looked to see if apparently more complex ones decomposed to these. The systems look different though they do the same things. It strikes us as quite possible that the CTM and MD-O are in a similar situation. At least at present we find no compelling way of distinguishing them.

- Bošković, Željko. 2007. On the locality and motivation of Move and Agree: An even more minimal theory. *Linguistic Inquiry* 38: 589-644.
- Chomsky, Noam. 1995. *The Minimalist Program*. MIT Press, Cambridge, Mass.
- Chomsky, N. 1993, A minimalist program for linguistic theory, in K. Hale & S. J. Keyser, eds, 'The view from Building 20', MIT Press, Cambridge, MA, pp. 1-52.
- Chomsky, Noam. 2008. On phases. in *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud*, eds. Robert Freidin, Carlos Otero, and Maria Luisa Zubizarreta, 133–166. Cambridge, Mass.: MIT Press.
- Chomsky, Noam and Lasnik, H. 1993. The theory of principles and parameters. In *Syntax: an international handbook of contemporary research*. von Stechow, J. Jacobs A., Sternefeld, W. & Vennemann, T. (eds.). Berlin: De Gruyter. Reprinted in: *The minimalist program, ch.1*. Chomsky 1995, 13-127. Cambridge MA: MIT Press.
- Epstein, Samuel David, Erich M. Groat, Ruriko Kawashima & Hisatsugu Kitahara. 1998.

- A Derivational Approach to Syntactic Relations. New York: Oxford University Press.
- Gärtner, Hans-Martin. 2002. *Generalized Transformations and Beyond: Reflections on Minimalist Syntax*. Berlin: Akademie Verlag.
- Hornstein, Norbert. 2001. *Move! A minimalist theory of construal*. Oxford: Blackwell.
- Hornstein, Norbert. 2009. *A theory of syntax: Minimal operations and universal grammar*. Cambridge: Cambridge University Press.
- Lebeaux, David. 1988. *Language acquisition and the form of the grammar*. Doctoral dissertation, University of Massachusetts, Amherst.
- Lebeaux, D. (1991). Relative Clauses, Licensing, and the Nature of the Derivation. In: Rothstein, S. D. (ed.) *Perspectives on Phrase Structure: Heads and Licensing*. San Diego: Academic Press
- Marcus, G. F. 2001. *The Algebraic Mind*. MIT Press, Cambridge, Massachusetts.
- Nunes, Jairo. 2004. *Linearization of Chains and Sideward Movement*. volume 43 of *Linguistic Inquiry Monographs*. MIT Press, Cambridge, Massachusetts.