Second Draft: October 2000
First Draft: August 2000
Minor revisions: February 2001

Antecedent Contained Deletion and The Copy Theory of Movement

Danny Fox, Harvard University

Structures with Antecedent Contained Deletion (ACD) have figured prominently in attempts to understand the mechanisms that account for the interpretation of quantifiers in natural language. The interest in these structures stems from the fact that they provide one of the strongest arguments that covert movement is a necessary piece of the account (Sag 1976, Larson and May 1990, and many others). The argument, however, turns out to conflict with a fairly well motivated idea regarding the nature of movement rules, namely the idea that movement is a copying operation which does not eliminate an element from its base position (the copy theory of movement).

My goal in this paper is to argue in favor of a specific resolution for the conflict. I begin by introducing some assumptions about the interpretation of chains and the structure of relative clauses. Once these assumptions are in place, the conflict is resolved with the aid of Lebeaux's (1988) proposal that movement can be followed by late insertion of an adjunct extended to covert movement by Fox and Nissenbaum (1999, in progress). In other words, the resolution turns out to have a familiar logic: the conflict is very similar to the conflict that Chomsky (1993) notes between the copy theory of movement and the observation that movement can sometimes obviate Condition C, and the resolution in both cases is basically identical. Furthermore, the resolution provides an account of a wide range of otherwise puzzling facts, thus supporting the assumptions that it utilizes as well as those that enter into the formulation of the conflict itself.

1. The Problem

It is well known that VP deletion is possible only if the elided VP bears some resemblance to an antecedent VP. I will state this parallelism requirement in the simplest possible way, ignoring interesting questions that arise from various counter-examples:[1]

(1) Parallelism: An elided VP must be identical to an antecedent VP at Logical Form (LF).[2]

---

[1] One (potential) counter-example, which bears on the issues discussed in this paper, has to do with the fact that some properties of DPs that are relevant for Binding Theory (BT) don't seem to have effects in the elided VP (Vehicle Change, VC). Brody (1995) and Wyngaerd and Zwart (1991) argue that VC resolves the conflict between the structure in (2) and Parallelism. However, this is not obvious. There are many ways to think of VC under which the conflict is not eliminated. It is possible, for example, that VC should be dealt with not by relaxing Parallelism, but rather by revising certain assumptions about the way BT should apply to elided material. (See Fox 1993 and Williams 1995 for proposals along this line.)

One problem with the idea that VC resolves the conflict is that (in and of itself) it does not account for Sag's correlation between ACD and Scope (section 5, but see Brody's proposal, which deals with this problem), as well as for additional evidence that ACD resolution requires QR (Pesetsky in press). My hope is that the results of this paper will provide further evidence for a definition of Parallelism that rules out LF structures such as (2).

[2] Throughout this paper, I will pretend that variable names don't matter for Parallelism. The truth is, however, that they do matter: variables that bear different names must be bound from parallel positions (see Rooth 1992 for an account of this generalization.) Nevertheless, ignoring variable names will not have harmful effects since the variables will always be bound from parallel positions (but see notes 28 and 29).

Parallelism can be used to investigate the nature of LF structures. In particular, ACD suggests that quantificational phrases are sometimes not interpreted in their base positions. To see this, consider the sentence in (2). If the universal quantificational phrase were interpreted in its base position, the antecedent VP would contain the elided VP, and consequently the two VPs would not be identical. This means that if Parallelism is to be satisfied, the object in sentences such as (2) cannot be interpreted in its base position.

Elided VP

(2) John [$_{VP}$ likes every boy Mary does <likes t>].

Antecedent VP

If, however, the object is interpreted in a VP external position, the structure satisfies Parallelism:

(3) [every boy Mary does **< likes t>**]

John **likes t**

This observation argues that there is a covert operation that can move a DP from its base position without affecting the phonology of a sentence (Quantifier Raising, QR). The argument has been reinforced by the corroboration of various intricate predictions regarding the scope of a DP that has to move for ACD resolution. (See Sag 1976, Larson and May 1990. See also section 5.1.)

The argument in my view is quite impressive. A syntactic operation, which is needed under what is arguably the simplest theory of semantics (of the type developed, for example, in Heim and Kratzer 1998), seems to provide a simple account of ACD which generates non-trivial yet correct predictions. However, the account of ACD (and consequently the argument) relies on the assumption that traces are fairly impoverished in their representations, and, as such, is in conflict with the copy theory of movement.[3] Under the copy theory of movement traces are copies of their antecedents, as in (4), and it is therefore far from obvious that (covert) movement can account for the fact that ACD constructions satisfy Parallelism.

(4) [every boy Mary does <likes t>]

John likes [every boy Mary likes t].[4]

To further appreciate the conflict, let's consider an argument for the copy theory of movement that relies on Condition C of the binding theory. The argument is based on the observation that certain types of movement do not affect Condition C, an observation that is mysterious if traces are informationally impoverished but is expected under the copy theory of movement.

In (5)a, for example, *wh*-movement removes the name *John* from the c-command domain of the pronoun *he*. Nevertheless, coreference between the two elements is

---

[3] This conflict was noted in Fox (1995). There the resolution was based on an economy condition that allowed the copy at the tail of the chain to be eliminated when needed for ACD resolution. (See, also, Fox 1999a; 2000, Sauerland 1998, and Merchant forthcoming.) The current paper demonstrates that the approach to extraposition developed in Fox and Nissenbaum 1999 allows us to eliminate the economy condition with some empirical advantages (see notes 8 and 36).

[4] The careful reader might wonder about the representation that *t* should get under the copy theory of movement. To answer this question certain assumptions about the structure of relative clauses need to be made precise. Specifically one needs to be precise about the nature of the phrase that undergoes *wh*-movement in relative clauses. See 1.3.2.

impossible (but see note 9). Similarly, in (5)b, QR can move the universal DP outside the c-command domain of the pronoun *him.* (Notice the availability of inverse scope in *Someone introduced John$_i$ to every friend of his$_i$.*) QR should, therefore, allow coreference between the pronoun and a name contained within the DP, but it doesn't.

(5) a. ??Guess [which friend of John's$_i$] he$_i$ visited t.

b. ??/*Someone introduced him$_i$ to every friend of John's$_i$.

[every friend of John's$_i$] Someone introduced him$_i$ to t.

Neither *wh*-movement nor QR can reverse the verdict of Condition C. This observation is mysterious if traces are informationally impoverished as in (6) but is expected under the copy theory of movement, which postulates the structures in (7).

(6) a. [which friend of John's$_i$] he$_i$ visited t

b. [every friend of John's$_i$] Someone introduced him$_i$ to t

(7) a. [which friend of John's$_i$] he$_i$ visited [which friend of John's$_i$]

b. [every friend of John's$_i$] Someone introduced him$_i$ to [every friend of John's$_i$]

The argument for the copy theory of movement is based on the observation that movement cannot change the verdict of Condition C. This argument is in conflict with the argument for movement based on ACD. The latter is based on the idea that movement can change the verdict of Parallelism and crucially depends on representing traces as simple variables (as in the representation in (3)). If traces are copies as in (4), it is not clear that the output of QR can yield structures that satisfy Parallelism.

Why are Condition C and Parallelism so different? Why is it the case that movement is incapable of obviating a violation of Condition C, and yet is capable of obviating a violation of Parallelism? If we adopt the copy theory of movement, we seem to get the right results for Condition C but the wrong results for Parallelism. If we adopt a theory of movement that postulates impoverished traces, we get the right results for Parallelism but the wrong results for Condition C. In this paper I attempt to provide a solution to this puzzle. The basic idea is that, contrary to appearance, Condition C and Parallelism are not different. Movement is a copying operation and as such is incapable of obviating a violation of Parallelism in the same way that it is incapable of obviating a violation of Condition C. The appearance that movement is obviating a violation of Parallelism is very similar to the appearance that movement in other cases (discussed in Lebeaux 1988) is capable of obviating a violation of Condition C. In both cases it is not movement per se that obviates the violation. Rather, it is movement that allows for the late insertion of an adjunct, thus circumventing the need to ever create a configuration in which the violation occurs. But before this idea can be presented in any detail something needs be said regarding the interpretation of the syntactic structures derived by movement (viewed as a copying operation).

2. The Interpretation of Chains

Chomsky (1995) points out that the copy theory of movement simplifies syntax in two ways. First, the theory eliminates the need to postulate new objects (traces) beyond run-of-the-mill lexical items. In this sense, it brings us closer to a view of syntax as a recursive procedure that does not access anything but the lexicon. Second, the copy theory turns movement into a simpler operation; it is practically identical to the elementary structure-building operation (*merge*), differing only in that it takes as input an object that has served as input for an earlier merger (*move* is *re-merge*).

How do the copy theory of movement and the traditional alternative compare with respect to semantic interpretation? The traditional alternative creates structures for which interpretable mechanisms have been postulated: traces are interpreted as bound variables and moved operators (second order predicates) are integrated into the structure (by function application) once the operation of lambda abstraction is postulated. (See, among others, Heim and Kratzer 1998.) Can the same mechanisms be used when the copy theory of movement is adopted? Probably so: the semantic component can treat lower copies as variables. But, the adoption of the copy theory of movement opens other possibilities as well. Specifically, if operator variable constructions are to be formed, the copy at the tail of the chain can be converted to an element that contains a variable in various ways. In Fox (1999a), I suggested that this copy is converted to a definite description yielding interpretations similar to those of the following paraphrases:[5]

(8) Which boy Mary visited which boy?

Paraphrase: Which is the boy, *x*, such that Mary visited **the boy x**?

(9) every boy A girl talked to every boy.

Paraphrase: For every boy, *x*, there is a girl who talked to **the boy x**.

The boy x in the paraphrase is modeled on definite descriptions in natural language such as *the man John*. I assume that these definite descriptions are interpreted by predicate modification of *man'* and λx(x=*John'*) with the resulting predicate serving as argument of

[5] The suggestion was based on a proposal made in a different context by Rullmann and Beck (1998). In the current paper I spell out my suggestion in greater detail, and the resulting proposal is simpler (for reasons mentioned in note 3).

the determiner. This means that two operations are needed to convert the copy at the tail of the chain to an interpretable object:[6]

(10) Trace Conversion:

a. Variable Insertion: (Det) Pred (Det) [Pred λy(y=x)]

b. Determiner Replacement: (Det) [Pred λy(y=x)] the [Pred λy(y=x)]

With Trace Conversion together with λ abstraction, the structures created by movement/re-merge are interpretable:[7]

---

[6] Irene Heim (p.c.) points out that the definite article is the only determiner that can appear in natural language in the environment ___ *NP a* where *a* denotes an element of type e. This observation, she suggests, might make it possible to derive Determiner Replacement as a necessary consequence of Variable Insertion.

[7] Trace Conversion yields the right interpretation for chains that are headed by DPs in which D is conservative. Because natural language determiners denote conservative two-place predicates, there are no empirical problems that arise from Trace Conversion. Furthermore, Trace Conversion might play a role in an account of conservativity. Assume that there was a determiner whose denotation *D* was not conservative. In most cases discussed in the literature, this means that figuring out the truth value of *D*(A,B) requires verifying membership in B for individuals that are not members of A. However, given the copy theory of movement and Trace Conversion, the (characteristic function of the) second argument of *D* is a partial function defined only for elements that are members of A (B:= λx.B(theAx) = λx:A(x).B(x)). It is reasonable to assume that this situation would yield systematic presupposition failure (of the sort we observe in *Most students don't have a car and every student drives his car to school*, Beaver 1992).

(11) Which boy Mary visited which boy? --Trace Conversion -->

which boy λx [Mary visited **the boy x**]?

(12) every boy A girl talked to every boy. --Trace Conversion -->

every boy λx [a girl talked to **the boy x**]?

Trace Conversion, which allows chains to be interpreted, affects the representation of traces. However, in and of itself, this does not change the consequences of the copy theory presented in 1.1. It is still predicted that movement will not obviate a violation of Condition C and Parallelism; we still get the right results for Condition C and the wrong results for Parallelism. To see this, consider the effects that Trace Conversion has on the structures in (7)b and (4). A copy of *John* is still present in the representation of the trace in (7)b and the violation of Condition C is still correctly predicted:

(13) [every friend of John's$_i$] Someone introduced him$_i$ to [every friend of John's$_i$].

[every friend of John's$_i$] λx Someone introduced him$_i$ to

[**the friend of John's$_i$ (identical to) x**]

Chierchia (1995) also suggests that there is a possible link between conservativity and the copy theory of movement.

For additional evidence in favor of Trace Conversion, see Fox (in progress) and Sauerland (in progress).

Similarly, a copy of the antecedent VP is still present in the representation of the trace in (4) and a violation of Parallelism is still wrongly predicted:[8]

(14) [every boy Mary does <likes t>]

John likes **[the boy Mary likes t (identical to) x]**

3. The Proposed Solution

My proposed solution for ACD will be based on the idea that relative clauses can be merged with an NP after a DP that contains the NP has been moved (late insertion). Furthermore, the solution will depend on a particular assumption regarding the representation and interpretation of relative clauses. These ingredients are spelled out in 3.1 and 3.2.

3.1. Late Insertion

Recall the observation that *wh*-movement and QR are incapable of obviating a violation of Condition C, which was accounted for by the copy theory of movement. It is well known that the observation doesn't always hold. Specifically, *wh*-movement can bleed Condition C as long as the Condition C violation is due to a name contained within an

[8] In Fox (1999a, 2000), developing ideas of Fox (1995), I proposed that grammar allows for a more radical reduction of traces than that which results from Trace Conversion. In addition, I proposed that reduction is restricted by a principle of economy (see notes 3 and 36).

adjunct to the *wh*-phrase (Freidin 1986). This is seen when we contrast (5) above (repeated as (15)a and (15)b) with (15)c. (See also van Riemsdijk and Willaims 1981.)[9]

(15) a. ??/*Someone introduced him_i to every friend of John's_i.
[every friend of John's_i] Someone introduced him_i to t.
b. ??/*Guess [which friend of John's_i] he_i visited t.
c. Guess [which man that John_i likes] he_i visited t.

In (15)c, in contrast to (15)a and (15)b, movement is capable of bleeding a Condition C violation. Is this compatible with the copy theory of movement? Chomsky (1993) suggests that it is, once one adopts Lebeaux's (1988) proposal that relative clauses, and more generally adjuncts, can be added to a structure countercyclically. The relative clause *that* $\textit{John}_i$ *likes* can be merged with the NP *man* after *wh*-movement, yielding the structure in (16)a, which (under the assumptions made in this paper) is converted by Trace Conversion (and λ abstraction) to the structure in (16)b:

(16) a. [which man that John_i likes] he_i visited which man.
b. [which man that John_i likes]λx. he_i visited the man x.

In order to account for the unacceptability of (15)b, Chomsky claims, again following Lebeaux, that complements cannot be added countercyclically due to the projection

---

[9] Not all speakers get a clear contrast between (15)b and (15)c. One possibility is that the variation in judgment reflects a variation in the analyses of certain constituents as complements or as adjuncts (see Schütze 1995, Dowty 2001, Partee and Borchev 1998). See Kuno (1997) for skepticism regarding the significance of the contrast in (15), and Safir (1999) for an argument that the contrast emerges in a domain where Kuno’s skepticism does not apply.

principle. Given the set of assumptions adopted here, this idea can also be stated in terms of properties of the final representation. Adding the complement *of John's*$_i$ in (15)b after movement would yield the structures in (17), which are uninterpretable (due to a violation of the theta criterion).[10] The relational noun *friend* in the base position is missing an argument, and the definite description is consequently uninterpretable.[11]

---

[10] This is an advantage since it eliminates the need to postulate the theta criterion as a constraint on derivations (i.e., the projection principle). The theta criterion (as a constraint on representations) follows from the assumption that LF structures need to be interpreted, and is thus a "bare output condition" (see Chomsky 1995, Heim and Kratzer 1998). In general, the proposal made in this paper does not depend on properties of the derivation and could probably be stated in terms of constraints on the final representations (see Brody 1995).

However, Uli Sauerland (p.c.) points out that the projection principle might be needed to rule out a derivation of (15)b in which the NP *friend of John's* is merged to the determiner *which* after the latter has undergone *wh*-movement. Alternatively, one could imagine that movement of the determiner without its complement would be ruled out by the locality conditions on Head Movement.

[11] Optional complements might appear problematic. However, it is not obvious that they raise problems that are unique to this context. That is to say, many things that one might postulate in order to reconcile optional complements with the theta criterion could be adopted here with no additional complications. Suppose, for example, that one adopts the assumption that optionality results from two distinct lexical items (one requiring a complement, type *<e,et>*, the other disallowing it, type *et*). The only way for the complement to be merged with an interpretable result is if the type *<e,et>* element is selected from the lexicon, and then late merger would be ruled out for the reasons specified in the text. What if one assumes that optionality results from an operation that shifts the type of an *<e,et>* element to *et*? Nothing changes if the operation

(17) [which friend of John$_i$'s] he$_i$ visited which friend –Trace Conversion-->

[which friend of John$_i$'s] λx. he$_i$ visited the friend x.

So late insertion of an adjunct derives a structure in which the adjunct is missing from the tail of the chain. This accounts for the circumvention of a Condition C violation in (15)c, and will play a central role in the proposal for ACD. But first we have to discuss the way late insertion might interact with covert movement.

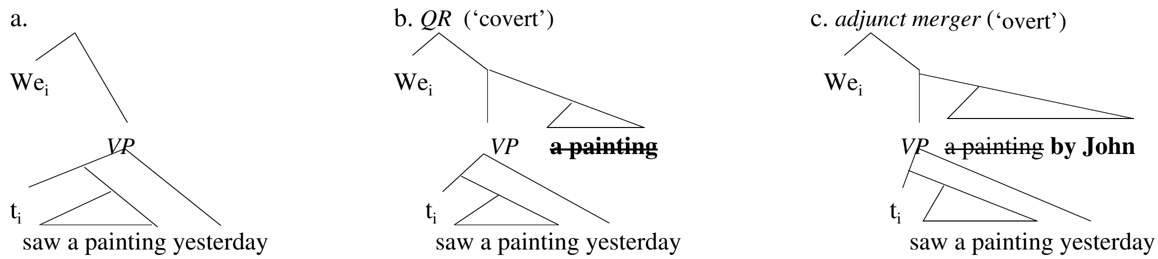
Under a model of grammar in which covert operations must follow all overt operations late insertion (an overt operation) cannot follow covert movement. However, Fox and Nissenbaum (1999, in progress) argue that late insertion can interact with covert movement in exactly the same way that it interacts with overt movement (though see Nissenbaum in progress for an important caveat). More specifically, they argue that extraposition of adjuncts from DP should be analyzed as involving QR of the DP followed by late insertion of an adjunct.

To understand the nature of the claim, consider the sentence in (18). Fox and Nissenbaum argue that this sentence is derived by covert QR followed by late insertion; the DP *a painting* (the DP from which, under traditional accounts, the adjunct has extraposed, henceforth, the source DP) undergoes QR to VP where the NP p*ainting* is later merged with the adjunct *by John*.

(18) We saw a painting yesterday by John.

---

applies in the lexicon. If the operation applies in the syntax, there will have to be a constraint that would rule out counter cyclic application of the operation at the tail of the chain.



[A painting by John] λx we saw [the painting x]

This order of operations is expected if covert and overt operations apply at the same cycle (as proposed in Brody 1997).[12] More specifically, it is a direct consequnce of a view of covert movement proposed by Bobaljik (1995, 1999), Groat and O'Neil (1996), and Pesetsky (1998), which we might call the phonological theory of QR. Under the phonological theory of QR, movement is sometimes covert not because phonology can pronounce pre-movement representations but rather because it can pronounce the copy at the tail of a chain. The correct word order in (18) is derived under the assumption that QR applies to the right: the source DP is pronounced in its base position because the movement is covert, and the adjunct is pronounced at the right periphery where it is merged.[13]

---

[12] After arguing for a single cycle, Brody moves on to argue that derivations should play no role in grammar (see note 10 as well as Brody 1995).

Brody's single cycle grammar has no covert *phrasal* movement, and (as pointed out by Bobaljik p.c.) is not obviously compatible wih the analysis that Fox and Nissenbaum provide for extraposition, at least not the way that the latter is stated.

[13] See Nissenbaum (in progress) for a view of grammar whereby the derivation in (18) does not depend on the phonological theory of QR. Nissenbaum presents intriguing arguments in favor of his alternative

Fox and Nissenbaum derive various predictions from this analysis, and I will present those that will be important for our discussion of ACD, namely those that have to do with scope and Condition C. The prediction for scope is straightforward. Because extraposition involves covert movement of the source DP to the position where the adjunct is merged, and because, due to the late merger, reconstruction is impossible, it is predicted that the scope of the source DP would be at least as high as the extraposition site.[14] That this prediction is correct was already noted in Williams (1974; chapter 4), and it is stated below as Williams' generalization. [15]

(19) **Williams' generalization:** When an adjunct β is extraposed from a source DP α, the scope of α is at least as high as the attachment site of β (the extraposition site).

---

proposal. However, the content of this paper does not depend on the choice between the two, and I will therefore keep to the more familiar proposal.

[14] If an adjunct is present only at the head of a chain, and if scope reconstruction results from interpreting only the tail of a chain (i.e., the head of the chain is deleted at LF), then late merger will block reconstruction; the adjunct would not be interpretable as a modifier of the source NP. Exactly these considerations are needed independently (as pointed out by Fox 1999a) to account for correlations between scope reconstruction and Condition C. Notice that in both cases there is an argument against "semantic reconstruction". If semantic reconstruction (via higher type traces) were possible, it would not conflict with late insertion of an adjunct.

[15] Williams, in contrast to Fox and Nissenbaum, focused on comparative- and result-extraposition, and not on extraposition from NP. It is also worth noting that he did not make the complement/adjunct distinction which is very central for Fox and Nissenbaum (see (27) below and note 18).

To illustrate the correctness of Williams' generalization I will make use of a scopal ambiguity discussed by Sag (1976) and Williams (1977) and illustrated by the sentences in (20). (For a more complete illustration of the generalization see Fox and Nissenbaum in progress.)

(20) a. I read a book before you did.
  b. I read every book before you did.

To understand the ambiguity, focus on (20)a. Under one interpretation, the sentence asserts that there exists a book such that the speaker read it before the addressee read it. Under another interpretation, the sentence asserts that the speaker read a book before the addressee read one. As both Sag and Williams argue, one interpretation results from a structure in which the indefinite *a book* is inside the antecedent for VP ellipsis, while the other interpretation results from a structure in which the indefinite is outside the antecedent VP and binds a variable in both the elided and antecedent VP.

The correctness of Williams' generalization is illustrated by the contrast in (21) and (22).

(21) a. I read a book that John had recommended before you did.
  b. I read a book before you did that John had recommended.

(22) a. I read every book that John had recommended before you did.
  b. I read every book before you did that John had recommended.

The (a) examples are ambiguous in the same way that the examples in (20) are. The (b) examples, by contrast, are restricted to the interpretation in which the source DP is outside the antecedent VP and binds a variable in the elided VP. The reason is straightforward. The extraposition site is above the *before*-clause, hence outside the antecedent VP. Consequently, by Williams' generalization, the source DP must be interpreted outside the antecedent VP. In terms of the theory, for the (b) sentences to be derived the source DP must move to a position above the *before*-clause where the relative clause can be merged, and therefore the source DP must have scope outside the antecedent VP.

Let us now move to the predictions that the analysis of extraposition has for Condition C. Because extraposition involves post QR insertion of adjuncts, it is predicted to have effects on Condition C of the binding theory (identical to the effects that were observed in the case of *wh*-movement, (15)c). Evidence that this prediction is correct is provided by the following contrast from Taraldsen (1981):

(23) a. I gave him_i a book yesterday that $John_i$ liked.

b. ??I gave him_i a book that $John_i$ liked yesterday.

Although the contrast is subtle (for reasons that will be addressed shortly), it seems that the indicated coreference is possible in (23)a, but slightly marked in (23)b. This is expected under Fox and Nissenbaum's approach to extraposition: in (23)a Condition C is obviated because the adjunct *that John liked* is merged with the source DP at the landing site of QR, yielding an interpretable structure in which the name *John* is not c-commanded by the coindexed pronoun:[16]

(24) I [[VP t gave him_i a book yesterday] ~~a book~~ that $John_i$ liked

[16] Notice, however, that it is not clear why (23)b cannot be derived in the same way (i.e., with string vacuous extraposition). Below, I suggest that it can, and that the contrast reflects a parsing preference that disfavors string vacuous extraposition (see p. 21).

[A book that John$_i$ likes] λx. I gave him$_i$ [the book x]

The standard explanation of Tarsldsen's contrast in (23) is that extraposition is a movement operation and as such is capable of bleeding Condition C. Evidence against the standard explanation and in favor of an explanation based on late insertion is provided by the contrast in (25).

(25) a. ? I told him$_i$ about your new argument the other day that supports John's$_i$ theory.

b. * I told you about his$_i$ new argument the other day that supports John's$_i$ theory.

In (25)a extraposition obviates a violation of Condition C.[17] This follows from the standard explanation as well as from the one based on late insertion. However, the fact that in (25)b extraposition does not obviate a violation of Condition C is extremely puzzling under the standard explanation; the pronoun *him*$_i$ in (25)a is in a higher (surface) position than the pronoun *his*$_i$ in (25)b; therefore, if extraposition is a movement operation capable of bleeding Condition C in (25)a, it should also be able to do so in (25)b.

On the late insertion account, the contrast in (25) is expected. If the effects on Condition C are due to late insertion, there are very specific predictions: a violation of the condition should be obviated only if the problematic pronoun (*him*$_i$ in(25)a and *his*$_i$ in

---

[17] The sentence is slightly marginal, but this is unrelated to Condition C: for reasons that are not clear to me, extraposition from a DP that has a possessor is not perfect. In any event, the contrast between (a) and (b) is pretty sharp.

(25)b) is not contained within the source DP. This is seen by the representations of the sentences in (25) that are derived by post QR insertion:

(26) a. I [[VP t told him$_i$ about your new argument the other day] ~~your new argument~~ that supports John's$_i$ theory

[your new argument that supports John's$_i$ theory] λx. I told him$_i$ about [the new argument x]

b. I [[VP t told you about his$_i$ new argument the other day] ~~his$_i$ new argument~~ that supports John's$_i$ theory

[his$_i$ new argument that supports John's$_i$ theory] λx. I told you about [the new argument x]

Additional evidence in favor of Fox and Nissenbaum's account comes from the observation that late insertion is possible for complements but not for adjuncts (see pp. 11-13). This observation leads to the prediction that complement extraposition will differ from adjunct extraposition in failing to bleed Condition C. The contrast in (27) suggests that the prediction is correct.

(27) a. I gave him$_i$ an argument yesterday **that supports John's$_i$ theory**.

b. ??/*I gave him$_i$ an argument yesterday **that John's$_i$ theory is correct**.

(27)a is derived by post QR insertion of the relative clause (in bold face), an operation which is expected to obviate Condition C. (27)b, by contrast, cannot be derived by late insertion of the bold faced phrase since this constituent is a complement. Instead, it is derived by rightward movement of a complement, which (given the copy theory of movement) does not obviate Condition C.[18]

Finally, the account in terms of late insertion predicts a correlation between the ability of extraposition to circumvent Condition C and the scope of the source DP. To understand this prediction, consider (28).

(28) a.? I wanted him_i not to talk to a (certain) girl yesterday that $John_i$ has known for years.[19]

b. * I wanted him_i not to talk to any girl yesterday that $John_i$ has known for years.

c. I wanted $John_i$ not to talk to any girl yesterday that he_i has known for years.

In (28)a, extraposition to the matrix clause is expected to yield the following structure in which Condition C is not violated:

---

[18] A derivation of adjunct extraposition via rightward movement needs to be blocked; otherwise the explanation Williams' generalization would be lost. This blocking is achieved in Fox and Nissenbaum (1999) with appeal to the observation that in general movement out of NPs is restricted to complements. See Fox and Nissenbaum (1999, in progress) for a variety of arguments that complement extraposition and adjunct extraposition have a very different analysis.

[19] (28)a is marginal for some speakers. However, it is definitely better than (28)b. The marginality of (28)a would follow if we assume that the parsing preference that disfavors string vacuous extraposition would extend in an appropriate way (see the discussion of (30) below). Specifically, we could assume that there is a preference for a structure in which the position of constituents is as low as possible (see Phillips 1996). Such a preference would favor a parse in which the relative clause is attached to the embedded VP (see note 37). (28)b is worse than (28)a since the former *requires* an unacceptable parse (while in the latter, this is only a preference).

(29) I [[$_{VP}$ t wanted him not to talk to a (certain) girl yesterday] ~~a (certain) girl~~ that John has known for years

[A (certain) girl that John has known for years]

λx. I wanted him not to talk to [the girl x] yesterday.

However, an obviation of Condition C by extraposition has predicted consequences for the scope of the source DP; it must be at least as high as the extraposition site. In the case of (28)a, the indefinite *a (certain) girl that John has known for years* must outscope both negation and the intensional verb *want*. This predicted consequence seems to be correct (as far as speakers can access the judgments). More importantly, the unacceptability of (28)b, verifies the prediction: the source DP is an NPI, which is not allowed to outscope negation, and thus brings about a conflict with the requirement imposed by Condition C. ((28)c is acceptable since Condition C imposes no requirements and extraposition to the embedded VP is available.)

Before we move on, it is important to comment on a question that arises from the contrast in (23). Why can't extraposition apply string vacuously in (23)b thus circumventing Condition C? The answer to this question (which arises for both the account provided by Taraldsen and for the alternative advocated here) is related, I would like to suggest, to the following three-way contrast:

(30) a. I gave him$_i$ a new argument the other day that supports John's$_i$ theory.

b. ??I gave him$_i$ a new argument that supports John's$_i$ theory.

c. *I gave him$_i$ a new argument that John's$_i$ theory is correct.

My bet is that string vacuous extraposition *is* possible but that it is disfavored given a parsing preference (perhaps a preference for low attachment of the sort argued for in Phillips, 1996). Specifically, I would like to suggest that the markedness of (23)b/(30)b is to be attributed to a parsing preference that chooses an analysis with no extraposition whenever possible. This preference is not very strong (perhaps even non-existent) for some speakers. However, for other speakers the preference has some effects and accounts for the contrast between (23)a/(30)a and (23)b/(30)b. (30)c is worse for all speakers. In this case the parsing preference is irrelevant; the name *John* is embedded in a complement to NP and counter-cyclic merger is impossible.

So scope and Condition C argue that it is possible for a covertly moved constituent to merge with an adjunct after movement has taken place. This possibility will explain the fact that a violation of Parallelism can be obviated despite the copy theory of movement. But first I will spell out my assumptions regarding the structure of relative clauses.

3.2. Relative Clauses

I will assume (following Sauerland 1998) that relative clauses are both head external and head internal (see also Cresti 1999 and Kennedy 2000). More specifically, I will assume that the derivation of relative clauses involves "movement to Comp" of a CP internal NP, which is deleted under identity with a CP external NP:[20]

---

[20] I cannot assume that relative clauses are totally head internal as was suggested in Vergnaud 1974. Vergnaud's suggestion is in conflict with the idea that relative clauses are adjuncts that can be generated independently of the head they modify and are thereby candidates for late insertion.

(31) Every boy [CP ~~boy~~ Mary likes ~~boy~~].

Furthermore, I will assume that the NP in [spec, CP] is not interpreted but that movement turns the relative clause into a predicate that combines with the CP external NP by predicate modification. More specifically, given Trace Conversion, we get the following structure:

(32) Every [boy λx. Mary likes the boy x]

Meaning: $\lambda P \forall x((boy(x)$ & Mary likes the boy $x) \rightarrow P(x))$

3.3. Explaining Antecedent Contained Deletion

If (32) is the interpreted structure of a DP containing a relative clause, and if the relative clause can be inserted into the structure after the DP has been shifted from its base position, then structures involving ACD can have the following derivation.

(33) [$_{VP}$ John likes every boy]. *–DP movement*--->

[[$_{VP}$ John likes] every boy] **every boy**. —*adjunct insertion*-->

[[$_{VP}$ John likes every boy] every boy **that Mary does <likes boy>**].

The first step of the derivation outlined in (33) is rightward movement of *every boy*. This movement is either overt (in which case the head of the chain is pronounced) or covert (in which case the tail is pronounced). In the former case the resolution of ACD involves

Heavy NP Shift (HNPS) and in the latter case it involves extraposition.[21] In both cases, Trace Conversion yields the structure in (34), which satisfies Parallelism.

(34) [every boy λx Mary does < likes the boy x>]
λy John likes the boy y]

Given the copy theory, ACD cannot be resolved by movement alone (whether it is overt or covert). However, movement allows for the late insertion of an adjunct that circumvents the effects that would otherwise result from the existence of the adjunct at the tail of the chain. Late insertion circumvents a violation of Parallelism in the case of ACD, just as it circumvents Condition C in the cases discussed by Lebeaux.

In the following sections, I will provide additional evidence in support of this proposal. In sections 4-6, I will show that certain phenomena, which were quite mysterious under the standard approach to ACD, are automatic consequences of this alternative proposal. In section 7, I will discuss Larson and May's criticism of a proposal made by Baltin (1987), which bears some similarity to the proposal made here. However, there are also important differences. It will turn out that due to the differences some of

[21] The idea that extraposition is necessary for ACD has been proposed in Baltin (1987) and has resurfaced in various forms (See Lasnik 1995, Wilder 1995a, Abe and Hoshi 1998). I share with Baltin and with the other researchers (besides Wilder, whose claim is more intricate) the idea that the relative clause containing the ellipsis site is outside the antecedent VP (and thus "antecedent contained deletion does not exist"). However, my proposal is (as far as I can see) the only one that derives the link with extraposition from independently motivated principles (see note 22 and section 7.1). Furthermore, it is the only one that provides a resolution to Kennedy's puzzle (section 5).

the criticism does not apply to the current proposal, and in section 8 I will argue that the portion which does apply turns out, on further scrutiny, to provide additional support.

4. Tiedeman's Puzzle

Larson and May (1990) claim that the unacceptability of the sentences in (35) is due to a locality condition on QR (Clause Boundedness).

(35) a. *I expected that everyone you do will visit Mary.
b. *I will say that everyone you will arrived.

However, this claim is challenged by the acceptability of the ACD constructions in (36), which were noted by Tiedeman (1995) and discussed extensively by Wilder (1995).

(36) a. I expect that everyone will visit Mary that you do.
b. I will say that everyone arrived that you will.

The contrast between (35) and (36) is quite puzzling under the standard approach to ACD.[22] However, it is a direct consequence of the proposal made here (once Clause Boundedness is abandoned).[23] Under the current proposal, for ACD to be resolved

---

[22] As far as I can see, all previous attempts to account for this contrast are based on the postulation of an ad hoc constraint. Wilder (1995a), for example, claims that the contrast argues that the ellipsis site cannot be PF-contained (a notion that he defines) within the antecedent VP. However, he does not derive this constraint from independent principles.

[23] For various independent challenges to clause boundedness, see May 1988, Reinhart 1991, and Wilder 1997.

movement to the right of a DP (overt or covert) must be followed by late insertion of the relative clause that contains ellipsis (RCE). In (35), in contrast to (36), word order indicates that the RCE has not been added after rightward movement.

To understand this explanation in further detail consider the derivation of the sentences in (36), focusing on (36)a:

(37) a. I expect that everyone will visit Mary. -- QR -->

b. I [[expect that everyone will visit Mary] ~~everyone~~]--adjunct insertion-->

c. I [[expect that everyone will visit Mary] ~~everyone~~ that you do <expect one will visit Mary>][24]

In the first step of the derivation, QR adjoins the quantifier *everyone* to the matrix VP. This movement is covert, which means that the copy at the tail of the chain is pronounced. After QR the RCE is inserted at the head of the chain. The resulting structure is converted (by Trace Conversion and λ abstraction) to the structure in (38), which satisfies Parallelism:

(38) Every [one λx. you <expect the one x will visit Mary>]

λy. I expect the one y will visit Mary.

[24] Wilder claims that the RCE in (36) is dominated by the antecedent VP. His claim is motivated by the Right Roof Constraint (RRC) which should block extraposition out of a tensed clause. I take the facts in (36) as arguments against the RRC. Furthermore, I (and my informants) disagree with Wilder that the following is unacceptable:

(i) I said that everyone would visit Mary when we talked that you did.

The sentences in (35), by contrast, cannot be derived in any way that would satisfy Parallelism. As was shown in the previous section, Parallelism can be satisfied only if the RCE is added after movement of the relevant quantifier. Because movement is to the right, the RCE cannot be followed by material from the antecedent VP.

The solution for Tiedeman's puzzle is straightforward under the approach to ACD advocated here. ACD is possible only if a DP is moved rightward out of the antecedent VP and only if the RCE is added after this movement. Whether the movement is overt or covert, it is predicted that the RCE will not be followed by material that belongs to the antecedent VP.[25] Hence the contrast between (35) and (36).[26]

5. Kennedy's Puzzle

Kennedy (1994) noticed a problem for the standard account of ACD which is exemplified by the sentences in (39) (see also Williams 1995 and Heim 1997).

(39) a. I saw a book about the man$_j$ you did <*saw t_j >

b. I like the car that belongs to the man$_j$ you do <*like t_j >

---

[25] As we will see in sections 7.2 and 8, this prediction needs to be qualified given the availability of rightward movement that can extract various constituent from within the antecedent VP. The explanation of Tiedeman's puzzle will not be affected by this qualification since a VP cannot be rightward shifted outside of a finite VP that contains it. (see 8.1.)

[26] One might wonder why we don't have a parallel derivation in which the embedded subject is overtly moved (HNPS). This question, for which I don't have an answer, arises independently of the proposal made here; subjects of finite clauses cannot undergo HNPS also in the absence of ACD.

In these sentences the elided VP cannot receive the interpretation indicated in the angular brackets (<,>). Thus, (39)a, if it is at all interpretable, cannot express the proposition that the man you saw is such that I saw a book about him.[27] Similarly, (39)b cannot be interpreted as a proposition that the car belonging to the man you like is liked by me. This is surprising given the standard theory of ACD. The standard theory derives the unavailable interpretation by QR of the matrix object, as shown in (40).

(40) a. [a book about the man$_j$ you <saw t_j>]$_i$

I saw t_i.

b. [The car that belongs to the man$_j$ you <like t_j>]$_i$

I like t_i

Under the standard procedure for the interpretation of chains, these structures would obey Parallelism. (See Heim 1997 for a detailed illustration of this point.)[28]

[27] The sentence might have an alternative (and irrelevant) interpretation in which the content of the elided VP is *saw a book about*. This interpretation (if it is available) results from a structure derived by QR of *the man* (followed by late insertion of an RCE).

[28] Notice that the variable names are not identical. This was pointed out in Kennedy 1994 and in Heim 1997. The question is, are the variables bound from parallel positions (see note 2)? Once the relevant notion of parallel binding is spelled out (Rooth's theory of ellipsis) Heim can present two possible sets of assumptions about the syntax semantics interface each of which yield different answers to this question: the standard set of assumptions, "the predicate hypothesis," yields a positive answer, whereas an alternative that Heim entertains, "the formula hypothesis," yields a negative answer. She presents Kennedy's observation as an argument in favor of the formula hypothesis, but, as Sauerland points out, the contrast in (42) might reverse the verdict (see note 29).

The unavailability of the intended interpretation has received much attention in the literature (Heim 1997, Jacobson 1998, Sauerland 1998). In this section I will present Sauerland's (1998) suggestion that the explanation of the phenomenon relies on the copy theory of movement. Although I believe the basic idea to be correct, Sauerland's implementation faces, as we will see, some difficulties that stem from his assumption that ACD resolution relies on QR of a DP that contains the RCE. This assumption, as pointed out in sections 1-2, is itself at odds with the copy theory of movement, a fact which leads Sauerland to a fairly complex set of assumptions about the interpretation of chains. Under the proposal made here there is no conflict between ACD and the copy theory of movement, and Sauerland's insight is captured with no ancillary complexities.

5.1. Sauerland's Proposal

Sauerland suggests that the unacceptability of (39) follows from the copy theory of movement and from the fact that this theory yields less impoverished traces than those in (40). More specifically, he assumes that the sentences in (39) have the following representation at the level where semantic interpretation applies:

(41) a. *[a book about the man_j you <saw man_j>]$_i$

I saw $book_i$.

b. *[The car that belongs to the man_j you <like man_j>]$_i$

I like car_i

Assuming that these are the correct structures, the antecedent and the elided VP are not identical, and the unavailability of the intended interpretation is derived from Parallelism.

As further evidence that the copy theory is involved in the explanation, Sauerland presents the contrast in (42).

(42) a. *I visited a city near the lake$_j$ John did <visited t$_j$>.

b. I visited a city near the city$_j$ John did <visited t$_j$>.[29]

This contrast, which does not follow under alternative approaches to the facts in (39), follows from Sauerland's approach, which assumes the structures in (43).[30]

(43) a. *[a city near the lake$_j$ John <visited lake$_j$>]$_i$

I visited city$_i$.

b. [a city near the city$_j$ John <visited city$_j$>]$_i$

I visited city$_i$.

---

[29] Sauerland (1998) has identified various speakers who claim that although (42)a is worse than (42)b, the latter is still worse than a standard ACD construction. If these judgments are correct, they can be accommodated by a theory that combines Heim's (1997) hypothesis about the way traces are interpreted (the "formula hypothesis") with the copy theory of movement and Trace Conversion. In the resulting theory, (42)a would involve two violations of Parallelism while (42)b would involve only one. However, see Sauerland (1998) for a simpler alternative based on the observation that that (42)b improves for the relevant speakers when the second instance of *city* is replaced by the anaphor *one*.

[30] Sauerland mentions that sentences such as those in (i) have the status of (42)a rather than that of (42)b. He shows that this is expected given the approach to Parallelism argued for in Fox (1999b).

(i) a. I visited a city near the place you did.

b. I visited a town near the city you did.

5.2. The Relevance of Late Insertion

The contrast in (42) is very instructive. It suggests the source of the problem in (39): the NP that restricts the quantifier that raises for ACD resolution (*book*, *car*) is not identical to the head that is modified by the RCE (*man*). Furthermore, the structures that Sauerland postulates explain why this identity is crucial. However, the principles of grammar that determine these structures are not obvious.

Sauerland suggests that the copy theory of movement is one of the principles involved. However, as we have already seen, the copy theory of movement is in conflict with the assumption that QR (alone) resolves ACD. In order to resolve this conflict Sauerland adopts an economy principle proposed by Fox (1995, 1999) along with some ancillary assumptions which together end up predicting the correct structures.[31]

For the purposes of this paper it is not necessary to go over Sauerland's proposal in detail. It is sufficient to note that the structures that Sauerland needs (or rather ones that are similar enough for the explanation to follow) are an automatic consequence of the approach to ACD developed in this paper. In other words, Fox's economy principle (which was itself designed to resolve the conflict between ACD and the copy theory of movement) as well as Sauerland's additional assumptions are no longer needed (see note 3).

---

[31] In addition to Fox's economy principle, Sauerland needs to assume that all DPs quantify over choice functions, thus ensuring that a copy will not be replaced by a simple variable in order to resolve ACD.

As outlined in section 3.3, garden variety ACD is derived in the following way:[32]

(44) $[_{VP}$ John likes every boy]. -*QR*--->

$[[_{VP}$ John likes every boy] ~~every boy~~]. –*adjunct insertion*-->

$[[_{VP}$ John likes every boy] ~~every boy~~ that Mary does <likes boy>].

[every boy λx Mary does < likes the boy x>]

λy John likes the boy y]

The final structure is acceptable since late insertion obviates a violation of Parallelism: the RCE has been added after movement, thereby allowing Trace Conversion to create identical definite descriptions in both the elided and antecedent VPs.

The explanation of Sauerland's contrast (between (42)a and (42)b) will be based on the observation that late insertion can obviate a violation of Parallelism in (42)b but not in (42)a. Let us start with the acceptable (42)b, which is derived in the following way:

(45) $[_{VP}$ I visited a city]. -*QR*--->

$[[_{VP}$ I visited a city] ~~a city~~]. –*adjunct insertion*-->

$[[_{VP}$ I visited a city] ~~a city~~ near the city that John did <visited city>].

[a city near the city λx John did <visited the city x>]

λy I visited the city y]

[32] For expository purposes I focus on the case in which DP movement is covert. The explanation is identical if the movement is overt; what is important is that the restrictor of the raised quantifier must be identical to the head internal to the RCE, modulo material in the late inserted adjunct.

This case is similar to the standard case in (44) in that late insertion of an adjunct allows for an obviation of Parallelism. But there is an important difference: the second step of the derivation inserts a larger adjunct in (45) than in (44). In (45) the adjunct is the PP headed by *near*, which contains the RCE, while in (44) the adjunct is smaller; it is the RCE alone. The general principle is that all material belonging to the restrictor of the QR-ed DP that is not identical to the head modified by the RCE must be added after QR. Otherwise Parallelism will not be satisfied.

Now let's move to the unacceptable (42)a. Let's start with the observation that a derivation parallel to the one in (45) would result in a structure that violates Parallelism:

(46) [$_{\text{VP}}$ I visited a city]. -*QR*--->

[[$_{\text{VP}}$ I visited a city] ~~a city~~]. –*adjunct insertion*-->

[[$_{\text{VP}}$ I visited a city] ~~a city~~ near the lake that John did <visited lake>].

[a city near the lake λx John did <visited the lake x>]

λy I visited the city y]

The final representation does not satisfy Parallelism because the definite article has a different restrictor in the antecedent than it does in the elided VP. More generally, in sentence (42)a (as well as in Kennedy's cases in (39)) there is no way for late insertion of an adjunct to obviate a violation of Parallelism. The reason is that the head of the NP that restricts that quantifier that undergoes QR is not identical to the head modified by the

RCE. Because the former is not an adjunct, it cannot be added after QR and a violation of Parallelism cannot be obviated.[33]

5.3. Further Evidence

We started this paper with an observation of a conflict between the standard account of ACD and the copy theory of movement. I suggested that the conflict is resolved by late insertion of the RCE. Kennedy's observation and in particular Sauerland's minimal pair in (42) support this suggestion. Specifically, both seem to indicate that when late insertion cannot resolve the conflict, ACD is impossible. We will now see further indication that this is the case.

5.3.1 Complements vs. Adjuncts: As mentioned in 3.1 adjuncts can be merged countercyclically but complements cannot. (Late insertion of complements yields uninterpretable structures, as illustrated in (17).) This predicts the contrast between (47) and (48) also noted by Sauerland.

---

[33] The explanation provided by Sauerland (1998) for Kennedy's puzzle appears to conflict with the availability of ellipsis in (i).

(i) a. This is the man that John likes and this is the woman that Bill does.

b. I know which man John likes and which woman Bill does.

This (apparent) conflict carries over to the account provided here. However, Sauerland demonstrates that the conflict can be eliminated with more precise assumptions about the nature of Parallelism and the way it relates to focus. See Sauerland (1998: chapter 3) and, in particular, his discussion of examples (71) and (72).

(47) a. I visited a city **near the city John did <visited>.**

b. I made an argument **that was very similar to the one/argument you did <made>.**

c. I read a book **that discusses the book you did <read>.**

(48) a. *I made an argument **that we should adopt the argument you did <made>.**

b. ??I read a book **about the book you did <read>.**

The sentences in (47) have derivations along the lines of (45) with the bold-faced adjunct added after QR. Parallel derivations are impossible for (48) since the bold-faced phrase is a complement and cannot be merged countercyclically.

Our approach to ACD predicts that exceptions to Kennedy's observation of the type noted by Sauerland will be restricted to cases in which the head of the NP modified by the RCE is contained within an adjunct, and the facts seem to conform to this prediction.

5.3.2. When late insertion is visible: The account of Sauerland's paradigm was based on the assumption that exceptions to Kennedy's observation require late insertion of a *particular* adjunct (the bold-faced phrase in (47)). This assumption can be tested in cases in which late insertion has visible effects. Consider from this perspective the contrast in(49).

(49) a. I visited a city near the city that John did.

b. I visited a city yesterday near the city that John did.

c. ??/* I visited a city near the city yesterday that John did.[34]

This contrast suggests that late insertion is relevant for the account of Sauerland's paradigm. In (49)a and (49)b the constituent appropriate for satisfying Parallelism has been added countercyclically. In (49)c, by contrast, the late inserted adjunct is too small for Parallelism to be satisfied.

To see this in greater detail, let's consider the derivations. The derivation of (49)a, a repetition of (42)b, was provided in (45). (49)b has a parallel derivation but in this case the effects of late insertion are visible given the presence of the adverb *yesterday*. As we can see, the final representation satisfies Parallelism.

(50) [[$_{VP}$ I visited a city] yesterday]. -*QR*--->

[[[$_{VP}$ I visited a city] yesterday] ~~a city~~]. –*adjunct insertion*-->

[[[$_{VP}$ I visited a city] yesterday] ~~a city~~ near the city that John did <visited city>].

[34] (49)c seems to be unacceptable independently of the presence of ACD, and therefore seems irrelevant for our purposes. However, this objection does not hold for the following paradigm, the appreciation of which requires understanding Saueralnd's explanation of the compatibility of (ia) with the discussion in 5.1. (see note 30.

(i) a. I visited a town near the city that John did.

b. I visited a town yesterday near the city that John did.

c. ??/* I visited a town near the city yesterday that John did.

d. ? I visited a town near the city yesterday that was described in the newspaper.

[a city near the city   λx [John did   <visited the city x>]]

λy [I   visited the city y] yesterday

Finally, consider (49)c in which a smaller constituent has been added countercyclically:

(51) [$_{VP}$ I visited a city near the city]. -*QR*--->

[[$_{VP}$ I visited a city near the city] ~~a city near the city~~]. –*adjunct insertion*-->

[[$_{VP}$ I visited a city near the city] ~~a city near the city~~ that John did <visited city>].

[a city near the city   λx John did   <visited the city near the city x>]

λy I   visited the city y]

This derivation yields a representation that doesn't satisfy Parallelism and therefore (49)c is unacceptable.[35]

6. Antecedent Contained Deletion and Condition C

Fiengo and May (1994) claim that instances of QR which are needed for ACD resolution have consequences for Condition C. This claim is based on the acceptability of sentences such as those in (52).

(52) a. You sent him$_i$ the letter that John$_i$ expected you would.

---

[35] We also need to consider a derivation of (49)c in which there are two instances of QR, each followed by late insertion: the matrix object undergoes string vacuous QR, then the PP is merged; this is followed by QR of the embedded definite phrase to the right of the adverb where the RCE is merged. Such a derivation should be ruled out by the CED.

b. You introduced him$_i$ to everyone John$_i$ wanted you to.

c. I reported him$_i$ to every cop John$_i$ was afraid I would.

In these sentences, ACD resolution requires that the DP (containing the deletion site) will undergo QR, and QR has consequences for Condition C. The obvious question that arises is why instances of QR needed for ACD resolution differ from normal instances of QR, which, as we saw, are incapable of affecting Condition C. (See, section 1 and in particular the discussion of (5)b.)

Fiengo and May have provided one answer to this question, and I tried to argue for an alternative (Fox 1995, 1999). However, both proposals are based on extraneous assumptions that can be eliminated once the approach to ACD advocated here is adopted.[36] Under the proposal made here it is not QR that obviates Condition C but rather late insertion, which is needed to satisfy Parallelism. Specifically, for the construction in (52) to satisfy Parallelism the RCE must be merged with an NP only after rightward movement. This sequence of operations yields the following structures, which do not violate Condition C:

(53) a. [the letter λx John$_i$ expected you would < send him$_i$ the letter x>]
λy you sent him$_i$ the letter y.

b. [every one λx John$_i$ wanted you to < introduce him$_i$ to the one x>]
λy you introduced him$_i$ to the one y.

[36] Fiengo and May propose a fairly intricate principle that determines the ordering of Binding Theory and QR. Fox (1995) proposes an economy condition on the elimination of material from the tail of chains. Both proposals are made to account for the correlation with Condition C, and are not supported independently.

b. [every cop λx John$_i$ was afraid I would < report him$_i$ to the cop x>]
λy you reported him$_i$ to the cop y.

Fiengo and May also report a contrast between sentences such as those in (52) and similar sentences that lack ACD:

(54) a. ?? You sent him$_i$ the letter that John$_i$ expected you would write.
b. ?? You introduced him$_i$ to everyone John$_i$ wanted you to meet.
c. ?? I reported him$_i$ to every cop John$_i$ was afraid of.

Specifically, they argue that QR can obviate Condition C only if ACD is involved. This judgment is a little subtle; all speakers seem to allow coreference in the sentences in (52) but they seem to differ on whether or not they allow coreference in (54).

I would like to suggest that the difference among speakers has to do with the effects of a general preference against vacuous extraposition. More specifically, I would like to suggest that the contrast between (52) and (54) is identical to Taraldsen's contrast in (23) repeated below:

(55) a. I gave him$_j$ a book yesterday that John$_j$ liked.
b. ??I gave him$_j$ a book that John$_j$ liked yesterday.

The parser prefers to parse relative clauses as sisters of the pronounced NPs that they modify. The grammar allows an alternative parse (via QR and late merger of a relative clause with an unpronounced NP) but the parser chooses this structure only when there is evidence that it is needed. In (55)a the relevant evidence comes from word order and in (52) it comes from ellipsis. In (55)b and (54) the parser has no evidence that extraposition is needed. Therefore, extraposition is not postulated and as a result Condition C is violated.[37]

Finally, the account of ACD in terms of late insertion captures the contrast in (56) noted independently by Sauerland (1998) and Merchant (in press):

(56) a. I asked him$_i$ to teach the book about Mary that John$_i$ wanted me to <ask him$_i$ to teach>.

b. *I asked him$_i$ to teach the book about John$_i$ that Mary wanted me to <ask him$_i$ to teach>.

ACD resolution in (56) requires that the RCE be inserted after QR, yielding the structures in (57):

---

[37] This analysis is similar to Fox's (1995); both postulate a preference principle that has (detectable) consequences for Condition C but not for Parallelism. See Phillips (1996) for other arguments that Condition C (in contrast to word order and Parallelism) does not affect the (detectable) preferences of the parser.

As an extension to Fiengo and May's paradigm, Fox (1995) observes that in (i) Condition C is circumvented only if the matrix VP is elided. This observation can be accommodated under the present account if one assumes that the parsing preference is for the lowest possible attachment for the relative clause (see note 19 as well as section 8.2).

(i) I wanted him to read the books that John thought I did.

a. <*read> b. <wanted him to read>.

(57) a. [the book about Mary λx $John_i$ wanted me to < ask him_i to teach the book about Mary x>]

λy I asked him_i to teach the book about Mary y.

b. [the book about $John_i$ λx Mary wanted me to < ask him_i to teach the book about $John_i$ x>]

λy I asked him_i to teach the book about $John_i$ y.

These structures explain why Condition C is not violated in (56)a. In (56)b, by contrast, late insertion of the RCE would not obviate Condition C, as we see in (57)b. Furthermore, the name *John* is contained within a complement and there is subsequently no option for late insertion that would obviate Condition C.[38]

7. Larson and May's Counter Arguments

In the previous sections I have argued that constructions with ACD satisfy Parallelism due to two operations: a DP moves out of the antecedent VP and the RCE is merged countercyclically. When the first operation is covert, the second operation yields extraposition. For this reason my proposal is somewhat similar to Baltin's proposal (1987) that ACD resolution involves extraposition.

---

[38] Merchant (in press) claims that the judgments remain the same when (56)b is changed so that the name *John* is contained within an adjunct:

(i) I reported her_i to every cop in Abby's$_i$ neighborhood you did.

I share this judgment but some of my informants do not. This variation might be related to the parsing preference that disfavors string vacuous extraposition.

Baltin's proposal has been criticized quite extensively by Larson and May (1990, L&M). The goal of this section is to examine the relevance of L&M's criticism for the current proposal. It will turn out that the points of criticism can be divided into two parts. One part, although quite relevant under the standard analyses of extraposition, is not relevant once the approach developed by Fox and Nissenbaum is adopted. The other part, which is still relevant, will force me to provide certain ACD constructions with a fairly baroque analysis. However, I will show that this move makes some new and non-trivial predictions. My attempt to corroborate the predictions, if successful, will provide additional evidence in favor of the proposal defended in this paper.

7.1. The No Longer Relevant Criticism

Baltin (1987) argues that antecedent containment does not exist since the putative ACD constructions involve (overt) extraposition. More specifically, he assumes that the RCE is shifted to the right thereby removing the elided VP from within the antecedent VP:

(58) I [$_{\text{Antecedent-VP}}$ read every book t_i] [that you did <read t_j>]$_i$

L&M point out that it is not clear why extraposition should be required. It is true that extraposition removes the elided VP from within the antecedent. However, it is not true that the challenge for the theory of ellipsis is eliminated along with this elimination of antecedent containment. Despite the lack of antecedent containment, the antecedent and the elided VP in (58) are not identical and the construction does not satisfy Parallelism. In other words, there is no independently motivated condition that forces rightward movement of the RCE. If ACD constructions involve extraposition, it is not clear why; extraposition does not play a role in the resolution of the problem that the construction brings up.[39]

This criticism is (of course) not relevant for the approach to ACD advocated here. Under the current approach extraposition does not involve movement of the RCE, but rather late insertion, and the necessity of late insertion follows from Parallelism given the copy theory of movement.

An additional point of criticism is that the alternative makes correct predictions that are not captured by Baltin. If ACD is resolved by QR, there are very clear consequences for scope: the QR-ed DP must outscope all of the elements that take their scope within the antecedent VP. This prediction has been pointed out and corroborated by Sag (1976), and has subsequently been discussed in much literature. Thus, for example, the DP that contains ellipsis must outscope the intensional verb *want* in (59)a, and *refuse* in (59)b.[40]

(59) a. I want a car that you do <want>.

b. I refused to read every book that you did <refused to read>.

---

[39] Of course one can assume QR along with an additional condition on ellipsis that would ban surface antecedent containment (cf. Wilder 1995a). The additional condition might motivate extraposition, but as long as it has no independent motivation, it amounts to a stipulation of extraposition (or so it seems to me).

[40] (59)a is modeled on examples from Sag (1976), and (59)b was brought up by Irene Heim (during a class on ellipsis co-taught with David Pesetsky).

Under Baltin's approach to extraposition, it is not obvious that this prediction follows. There is no reason why rightward movement of the RCE will have consequences for the scope of the DP that contains the RCE.[41]

Once again, this criticism does not apply to the approach to ACD argued for here, since under this approach extraposition is not an alternative to QR. Extraposition is needed because (assuming the copy theory of movement) QR is not enough. Extraposition of a relative clause is itself an indication that QR has taken place, and the consequences for scope are retained.

L&M's third point of criticism relates to a potential diagnostic of extraposition: the complementizer *that* cannot be omitted in an extraposed relative clause. This diagnostic, which is supported by the unacceptability of omission in (60)a, suggests, as (60)b shows, that extraposition is not obligatory in ACD constructions.

(60) a. I read something yesterday *(that) John had recommended.

b. I read something (that) John did <read>.

Since we don't understand the principles that govern *that* omission, this argument is not as strong as the previous arguments. More specifically, there is a way to think of (60)a which will not be problematic for the assumption that there is extraposition in (60)b. It is not impossible that *that* omission is restricted by phonology rather than by

[41] Baltin (p.c.) points out that the prediction is captured by his analysis if one adopted Geuron and May's (1984) theory of extraposition in which QR is needed in order for extraposed constructions to receive an interpretation. This, however, makes Larson and May's first criticism more acute. If one assumes that QR takes place in ACD constructions, why is extraposition needed? The copy theory of movement, when accompanied by the assumptions made here, provides an answer.

syntax. More specifically, it is possible that the head of a relative clause cannot be omitted when the relative clause is not string adjacent to the NP that the relative clause modifies. If this is the constraint that accounts for (60)a, there will be no ban on *that* omission when extraposition is string-vacuous, hence no conflict between Baltin's approach to ACD and the availability of *that* omission in (60)b.[42]

Furthermore, even if the ban on *that* omission were a diagnostic of extraposition, it would not rule out the analysis of ACD proposed here. Under the analysis proposed here, late insertion of the RCE is obligatory for ACD resolution. But late insertion results in extraposition only when movement of the relevant DP is covert. When movement is overt (when ACD is resolved by HNPS), *that* omission will not be banned, thus accounting for the optionality observed in (60)b.[43]

7.2. The Still Relevant Criticism

If ACD requires extraposition, it is predicted that the RCE will be outside the antecedent VP. However, as pointed out by L&M, the acceptability of the ACD constructions in (61) suggests that this is not always the case. In (61) the RCE (in bold face) is followed by material that belongs to the antecedent VP, and is therefore plausibly contained within the antecedent VP.

(61) a. I [gave a book on linguistics **that you did <give t to Mary>** to Mary].

---

[42] Of course the phonological constraint has no explanatory power. Still, for L&M's argument to have much force one would want to show that the alternative (syntactic) constraint is better.

[43] Unfortunately, I can't think of any properties that might distinguish HNPS from extraposition and subsequently lead to interesting predictions.

b. I [gave everyone **you did <give a book on linguistics>** a book on linguistics].

c. I [wanted the man **you did <want to come to the party>** to come to the party].

If indeed the RCE is contained within the antecedent VP in (61), then extraposition is not involved in the derivation of these constructions. This criticism of Baltin's proposal extends to the proposal defended here. Under this proposal the RCE is merged after rightward movement of a DP. Whether the movement is overt or covert, the RCE must be outside of the antecedent VP.

However, there is a possible response to the criticism: the constituent that follows the RCE in (61) might also be outside of the antecedent VP (due to the availability of rightward movement). That this is possible is suggested by the acceptability of the sentences in (62). These sentences seem to involve rightward movement of the constituent that follows the RCE in (61).[44]

---

[44] A by now familiar analysis of (62)c involves leftward movement of the ECM subject ("raising to object"). If this analysis is correct, (61)c doesn't challenge the account of ACD. (ACD would be resolved by object shift followed by late insertion of the RCE.) As pointed out to me by Chris Kennedy, (i) is a more challenging example. Restructuring not withstanding, (i) would probably require an analysis of (62)c in terms of rightward movement of the RCE. Similar consequences would follow if (ii) is acceptable (cf. Kennedy 1997):

(i) I wanted to expect everyone you did <wanted to expect to come to the party> to come to the party.

(ii) ?? I wanted a picture of everyone you did <wanted a picture of to be on sale> to be on sale.

(62) a. I gave a book yesterday to Mary.

b. I gave everyone yesterday a book linguistics.

c. I wanted this man with all my heart to come to the party.

Given the availability of rightward movement, the sentences in (61) can be derived in the following way:

(63) [$_{VP}$ I gave a book to Mary] –*QR* or *HNPS*--->

[[$_{VP}$ I gave a book to Mary] a book] –*adjunct insertion*-->

[[$_{VP}$ I gave a book to Mary] a book you did <give to Mary>]

–*additional rightward movement*-->

[[[$_{VP}$ I gave a book ~~to Mary~~] a book you did <give to Mary>] to Mary][45]

This derivation is admittedly quite baroque. However, if we can find evidence in its favor, we will have additional support for the account of ACD.

Put differently, the account of ACD advocated in this paper makes a precise prediction about "sandwiched" ACD constructions in which material from the antecedent VP follows the RCE (SACD constructions):

(64) In SACD constructions, the post RCE material belonging to the antecedent VP (henceforth, PRCE) is shifted rightward outside of the antecedent VP.

(cf. Wilder 1995a)

---

[45] As we see from the final representation, the same word order is derived whether or not the head or the tail of *a book* is pronounced (whether ACD resolution involves QR or HNPS). The analysis predicts both to be possible.

If this prediction can be corroborated, we will have further evidence in favor of the account.

8. Further Evidence

In this section I will try to identify three diagnostics of rightward movement and use them to corroborate the prediction in (64). The first diagnostic is a constraint limiting movement of predicates needed for a full account of the facts discussed in section 4, the second diagnostic is Condition C, and the third diagnostic relates to the interpretation of constructions that involve movement of reason and manner adverbs (Stroik 1992).[46]

---

[46] David Pesetsky (p.c.) and Lisa Travis (p.c.) mention the contrast between (i) and (ii) as potential evidence against (64).

(i) * Which boy did you give a book yesterday to t?

(ii) ? Which boy did you give a book Mary did to t?

(iii) Which boy did you give a book yesterday to a friend of t?

The unacceptability of (i) suggests that there might be a constraint blocking movement out of rightward shifted constituents (perhaps a consequence of the CED), and this constraints seems to indicate that in (ii) the PRCE need not be rightward shifted. However, the status of (iii) suggests that the situation is more complicated. All I can appeal to at the moment is a hope that the correct explanation of the paradigm will not conflict with (64).

8.1. Rightward Movement of Predicates

An embedded (finite) VP cannot shift rightward out of a VP that dominates it. This is exemplified by the fact that a matrix adverb cannot intervene between an embedded finite VP and its subject:

(65) a. *I hope that Bill with all my heart will come to the party.
     b. *I told you that Bill when we met will come to the party.

It is, consequently, predicted that the matrix VP cannot be an antecedent VP in an SACD construction where the embedded VP serves as the PRCE. That this prediction is correct is exemplified by Tiedeman's examples discussed in section 4 and repeated below:

(66) a. *I expected that everyone you do will visit Mary.
     b. I expect that everyone will visit Mary that you do.
     (Tiedeman 1995, Wilder 1995a)

(67) a. *John believed that everyone you did was a genius.
     b. John believed that everyone was a genius that you did.

In the (a) examples the embedded VP serves as the PRCE. Because the PRCE cannot be shifted out of the antecedent VP, the examples are predicted to be unacceptable.

Similar evidence that (64) is correct was pointed out by Baltin (p.c.). Embedded infinitivals headed by the complementizer *for* are similar to embedded finite clauses in that the post subject phrase headed by *to* cannot be shifted out of the matrix VP:

(68) *I want for Bill with all my heart to come to the party.

We thus predict the following contrast:[47]

(69) a. *John wants for everyone you do to have fun.

b. John wants for everyone to have fun that you do.

(Baltin, p.c.)

8.2. Condition C

There are also potential consequences of rightward movement for Condition C. To see this consider the contrast between (70)a and (70)b.

(70) a. ??I asked him$_1$ to give a book [to someone in John's$_1$ building] when I met him$_1$.

b. I asked him$_1$ to give a book t$_2$ when I met him$_1$ [to someone in John's$_1$ building]$_2$.

This contrast is expected given the fact that in (70)a the dative argument is rightward shifted out of the c-command domain of the matrix pronoun *him* (and given the parsing preference for low attachment (see notes 19 and 37)).[48] Furthermore, the contrast suggests a potential test for the prediction that the PRCE is rightward shifted.

[47] For an additional contrast that might fall under (64) see Jacobson (1992:162, example (37))

[48] Rember that the contrast should parallel the contrast in (30) above, and should therefore not hold for all speakers.

Consider from this prespective the contrast in (71). In (71)a, there is no preasure to parse the dative argument as a rightward shifted constituent. Consequently, the sentence violates Condition C. (71)b, by contrast, is an SACD construction with the dative argument serving as the PRCE. Given (64), the dative argument must be rightward shifted, and Condition C is subsequently obviated.

(71) a. ??I asked him_1 to give the book that you read [to someone in $John's_1$ building].

b. I asked him_1 to give the book t_2 that you did <ask him_1 to give t_2> [to someone in $John's_1$ building]$_2$.

Finally, consider the contrast in (72). (72)a is an SACD construction just like (72)b (= (71)b). However, in (72)a ACD can be resolved with the PRCE adjoining to the embedded VP. Consequently, given the parsing preference for low attachment, Condition C is not obviated. In (72)b, by contrast, the matrix VP is the antecedent for ellipsis, and (64) forces the PRCE to be shifted out of this VP thereby obviating Condition C.

(72) a. ?? I asked him_1 to give the book t_2 that you did <give t_2> [to someone in $John's_1$ building]$_2$.

b. I asked him_1 to give the book t_2 that you did <ask him_1 to give t_2> [to someone in $John's_1$ building]$_2$.

8.3. Movement of Reason and Manner Adverbs

Consider the contrast between the sentences in (73) and (74).

(73) a. I studied nothing in order to please the teacher that you did <study in order to please the teacher>.

b. Bill read every book with great care that I did <read with great care>.

(74) a. I studied nothing that you did <*study in order to please the teacher> in order to please the teacher.

b. Bill read every book that I did <*read with great care> with great care.[49]

There are a few differences between the sentences. One difference which is not relevant for our purposes but which we should nevertheless be aware of is that in (74) the pronounced adverb can be parsed as part of the RCE. In (73) such a parse is impossible for obvious reasons.[50] The contrast I would like to focus on is different. It has to do with

---

[49] Jon Nissenbaum (p.c.) suggests that the contrast between (73)a and (74)a can be made clearer when we look at the following sentences in an appropriate context. In particular, he suggests that we consider a context in which the speaker is a diver in the Olympics who was told by her coach to do certain fancy dives in order to prove to the judges that she could do them (see also note 51):

(i) I did none of the dives in order to prove that I could that my coach told me to <do in order to prove that I could>.

(ii) I did none of the dives that my coach told me to <*do in order to prove that I could> in order to prove that I could.

[50] Another difference (which is also irrelevant in this context) has to do with the scope of the reason adverb in the (a) sentences. In (73)a the adverb must be interpreted within the scope of the negative quantifier. In (74)a the relative scope of the two elements is free. This contrast is an instance of Williams' generalization discussed in section 3.1.

the content of the elided VP when the pronounced adverb modifies the antecedent VP (a parse which, again, is necessary for (73) but optional for (74)). Under such a parse the elided VP can contain a copy of the adverb in (73) but not in (74).

To see this, let's focus on the (a) sentences. (73)a, when uttered by a speaker *s* to an addressee *a*, can assert the proposition that there is nothing that *a* studied in order to please the teacher such that *s* studied it in order to please the teacher; call this proposition *P*. (74)a cannot express *P*; when the pronounced adverb modifies the antecedent VP, the elided VP cannot contain a copy of the adverb (or a variable bound by that adverb).

This seems fairly clear to the speakers I consulted.[51] Verifying this based on judgments of assent requires us to consider a situation in which *P* is true while the propositions expressible by (74)a are all false.[52] Assume that *s* and *a* have studied the same four topics X-W, but for different reasons; *a* studied X and Y in order to please the teacher, and *s* studied these topics out of genuine interest; conversely *s* studied W and Z in order to please the teacher, and *a* studied these topics for other reasons:

---

[51] The contrast can be sharpened if a bound pronoun is embedded within the adverb. The fact that sloppy identity is possible when the adverb has a copy in the elided VP distinguishes the relevant readings in a fairl clear way:

(i) a. $John_i$ discussed nothing in order to please his_i teachers that $Bill_j$ did <discussed in order to pleas his_j teachers>.

b. $John_i$ discussed nothing that $Bill_j$ did <*discussed in order to pleas his_j teachers> in order to please his_i teachers.

[52] (74)a has (potentially) four interpretations: the pronounced adverb can be parsed in the matrix or within the RCE, and the relative scope of the negative quantifier and the adverb is not fixed (note 50). All of these interpretations must be false in the situation that serves as a test case.

	Studied in order to upset the teacher	Studied for other reasons
Studied by *a*	X, Y	W, Z
Studied by *s*	W, Z	X, Y

In this situation, there isn't any topic studied by *a* in order to please the teacher such that it has been studied by *s* in order to please the teacher. In other words *P* is true. However, it is a fairly clear judgment that (74)a, in contrast to (73)a, cannot be true in the situation. We conclude that (73)a and (74)a differ in that only the former can express *P*.

The difference between (73) and (74) follows from our proposal if there is a constraint that disallowed rightward movement of the adverbs in these sentences. The interpretation of (74) in which the adverb modifies both the antecedent and the elided VP could be derived only if rightward movement of the adverb (the PRCE) were possible (64). The putative constraint blocks rightward movement and explains the unavailability of the interpretation.

This means that if the constraint existed, the unavailability of the relevant interpretation would be a verifying instance of the prediction in (64). However, we would like to have independent evidence for the constraint. I will not attempt to provide this evidence. Instead, I would like to argue that the unavailability of the relevant interpretation can serve to verify (64) even if the evidence is unavailable. The argument is based on Stroik's (1992) observation that (leftward) movement of the relevant adverbs blocks the intended interpretation.

Consider the sentences in (75). These sentences are similar to the sentences in (74), and different from the sentences in (73), in that the adverb cannot modify the elided VP (when it also modifies the antecedent VP).[53]

(75) a. In order to please the teacher, I discussed nothing that you did.

1. <discussed t>

2. <*discussed t for that reason/ in order to please the teacher>

b. With great care, Bill read every book that I did.

1. <read t>

2. <*read t in that way>

The account of Stroik's observation is not very important in this context. (See Bissell 1999 for a proposal.) What is important for our purposes is that the unavailability of the intended interpretation of (74) follows under our proposal whether or not reason and manner adverbs can shift to the right. If there is a constraint that bans the adverbs from shifting to the right, the facts follow given the constraint, and if there is no such constraint, they follow as an instance of Stroik's generalization. In any event, the prediction in (64) is verified.

---

[53] Again, there are irrelevant properties that distinguish among the sentences and could bring about confusion. (75) is similar to (73) and different from (74) in that (for obvious reasons) it doesn't have the irrelevant parse in which the adverb modifies only the elided VP. Also, (75) is different from both (74) and (73) in that the adverb must outscope the negative quantifier (probably due to the fact that the negative quantifier is a weak island, cf. Frampton 1991, Cresti 1995).

9. Conclusions

Baltin (1987) proposed that ACD is resolved by extraposition. At the time, there were serious problems with the proposal. The conflict between ACD constructions and the theory of ellipsis was not resolved by extraposition and there was therefore no reason to assume that extraposition was involved. Furthermore Sag's correlation (between ACD and scope) was not accounted for.

This paper can be read as a revival of Baltin's proposal. With certain assumptions about the structure and interpretation of chains, extraposition, viewed as late insertion, resolves a conflict between the copy theory of movement and Parallelism. This is the theoretical reason to assume extraposition (or HNPS when QR is overt). The correlation with scope is, of course, no surprise.

The resulting analysis has some non-trivial advantages. It accounts for Kennedy's puzzle and Sauerland's extension to this puzzle, for the correlation between ACD and Condition C, and for the fact that constituents from the antecedent VP that follow the elided VP behave like (rightward) shifted constituents.

These advantages, in turn, provide evidence for the assumptions that underlie the analysis. We thus have new evidence for the copy theory of movement (Chomsky 1993), for counter-cyclic merger of adjuncts (Lebeaux 1988), for the extension of Lebeaux's idea to covert movement (Fox and Nissenbaum 1999, in progress), and for the idea that the copies at the tail of the chain should be interpreted as definite descriptions that contain variables (Fox 1999a).

References

Abe, Jun and Hiroto Hoshi (1999) "A generalized rightward movement analysis of Antecedent Contained Deletion," *Journal of Linguistics* 35, 451-487.

Beaver (1995) *Presupposition and Assertion in Dynamic Semantics*, Ph.D Dissertation, University of Edinburgh.

Bobaljik, Jonathan (1995) *Morphosyntax: the Syntax of Verbal Inflection* Doctoral dissertation, MIT, MITWPL.

Bobaljik, Jonathan (1999) "A-chains at the Interface: Copies, Agreement and 'Covert Movemetn'," unpublished ms. McGill University.

Brody, Michael (1995) *Lexico-Logical Form,* MIT Press.

Brody, Michael (1997) "Perfect Chains," in Haegeman, Liliane (ed.) *Elements of Grammar: Handbook in Generative Syntax (Kluwer International Handbooks of Linguistics,Vol 1)*, Kluwer.

Chierchia, Gennaro (1995) Talk given at Utrecht University.

Chomsky, Noam (1993) "A Minimalist Program for Linguistic Theory," in *The View From Building 20*, Cambridge, MA: MIT Press (re-printed as chapter 3 of Chomsky 1995)

Chomsky, Noam (1995) *The Minimalist Progrm*, Cambridge, MA: MIT Press.

Cresti, Diana (1995) "Extraction and Reconstruction," *Natural Language Semantics*,

Cresti, Diana (1999) "Ellipsis and Reconstruction in Relative Clauses," in *Proceedings of NELS 30*, GLSA Publications, Amherst.

Cullicover and Rochemont (1990) *English Focus Constructions and the Theory of Grammar*, Cambridge: Cambridge University Press.

Fiengo, Robert and Robert May (1994) *Indices and Identitu*, MIT Press.

Fox, Danny (1993) Chain and binding—A modification of Reinhart and Reuland's 'Reflexivity'. Ms., MIT.

Fox, Danny (1995) Condition C effects in ACD. In R. Pensalfini and H. Ura (eds.) *Chapters on minimalist syntax*, MITWPL 27. Cambridge, MA: MIT Working Chapters in Linguistics, pp. 105–119.

Fox, Danny (1999a) "Reconstruction, Binding Theory and the Interpretation of Chains," *Linguistic Inquiry*.

Fox, Danny (1999b) "Focus, Parallelism and Accommodation," *Proceedings of SALT 9*.

Fox, Danny (2000) *Economy and Semantic Interpretation*, MIT Press.

Fox, Danny (in progress) "Extraposition and Intensional Adjectives," Harvard University.

Fox, Danny and Jon Nissenbaum (1999) "Extraposition and Scope: A case for overt QR," WCCFL 18.

Fox, Danny and Jon Nissenbaum (in progress) "Extraposition and the Nature of Covert Movement," ms. Harvard University.

Frampton, John (1991) "Relativized Minimality: A review," *The Linguistic Review*, 8:1-46.

Friedin, Robert (1986) "Fundamental Issues in the Theory of Binding," in B. Lust, ed. *Studies in the Acquistion of Anaphora*, Dordrecht: Reidel.

Géuron Jacqueline and Robert May (1984). "Extraposition and Logical Form," *Linguistics Inquiry*, 7(3), 429-462.

Groat, Erich, and John O'Neil (1996) "Spellout at the LF interface," In Werner Abraham, Samuel David Epstein, H_skuldur Thr_insson & C. Jan-Wouter Zwart (eds.). Minimal Ideas: Syntactic Studies in the Minimalist Framework. John Benjamins Publishing Company, Amsterdam/Philadelphia, 113-139.

Heim, Irene (1997) "Predicates or Formulas? Evidence frem Ellipsis," In *Proceedings of SALT 7*, ed. By Aaron Lawon and Eund Cho, 197-221, Ithaca, New York: CLC Publications.

Heim, Irene and Angelika Kratzer (1998), *Semantics in Generative Grammar*, Blackwell, Oxford.

Jacobson, Pauline (1992) "Comment Flexible Categorial Grammar, Questions and Prospects," in Robert Levine (ed.), Formal Grammar, Theory and Interpretation, Oxford University Press.

Kayne, Richard (1998). "Overt vs. Covert Movements," *Syntax* 1:128-191.

Kennedy, Christopher, (1994). "Argument contained Ellipsis," Linguistics Research Center Report LKC-94-03, University of California, Santa Cruz.

Kennedy, Christopher, (1997). "Antecedent Contained Deletion and the Syntax of Quantifiers," *linguistic Inquiry*.

Kennedy, Christopher, (2000). "Comparative Deletion and Optimality in Syntax," Unpublished ms. Northwestern University.

Kuno, Susumu (1997) "Binding Theory and the Minimalist Program," Ms. Harvard University.

Lasnik, Howard (1995). "A Notve on Psuedogapping," In Rob Pensalfini and Hiroyuki Ura, eds. *Papers on Minimalist syntax*, 143-164. (MIT Working Papers in Linguistics 27.) MITWPL, Department of Linguistics and Philosophy, MIT, Cambridge, MA.

Larson, Richard and Robert May (1990). "Antecedent Containment or Vacuous Movement: Reply to Baltin," *Linguistic Inquiry*, 21, 103-122.

Lebeaux, David (1988). "Language Acquisition and the Form of the Grammar," Doctoral dissertation, University of Massachusetts, Amherst.

May, Robert (1988) "Ambiguities of Quantification," *LI*, 19:118-135.

Merchant, Jason (forthcoming) "On the Extent of Trace Deletion in ACD," to appear in *Linguistic Inquiry*.

Pesetsky, David, (1998) "Some Optimality Principles of Sentence Pronunciation," in P. Barbosa, et.al, eds. *Is the Best Good Enough? Optimality and Competition in Syntax*, MIT Press and MITWPL.

Pesetsky, David (in press) *Phrasal Movement and its Kin*, MIT Press.

Phillips, Colin (1996) *Order and Structure*, Doctoral dissertation, MIT, Cambridge, MA: MITWPL.

Reinhart, Tanya (1991) "Non-Quantificational LF," In A. Kasher (ed.) *The Chomskian turn*,

Riemsdijk, Henk van, and Edwin Williams (1981). "NP-Structure, *The Linguistic Review* 1:171-217.

Rooth, Mats (1992) "Ellipsis reduandancy and Reduction Redundancy," In S. Berman and A. Hestvik, eds., *Proceedings of Stutgart Ellipsis Workshop.* Arbeitspapiere des Sonderforschungsbereichs 340, Bericht Nr. 29. IBM Germany, Heidelberg.

Rullman, Hotze and Sigrid Beck (1997). "Reconstruction and the Interpretation of *Which*-Phrases," to appear in the proceedings of the Tüebingen Reconstruction Workshop.

Safir, Ken (1999) "Vehicle Change and Reconstruction of A-bar-Chains," *Linguistic Inquiry*, 30(4): 587-620.

Sag, Ivan (1976) *Deletion and Logical Form*, Doctoral dissertation, MIT, Cambridge, MA: MITWPL.

Sauerland, Uli (1998) *The Making and Meaning of Chains*, Doctoral dissertation, MIT, Cambridge, MA: MITWPL.

Sauerland, Uli (in progress) "A Case of Focus Driven QR in English, " university of Tuebingen.

Stroik, Thomas (1992). "Adverbs and Antecedent Contained Deletion," Minimalism, Scope and VP Structure.

Taraldsen, Tarald, (1981) "The Theoretical Interpretation of a Class of Marked Extractions," In A. Belletic, L. Brandi , and L. Rizzi (eds.), *Theory of Markedness* *in Generative Grammar, Proceedings of the 1979 Glow Conference*, 475-516. Scuola Normale Superiore, Pisa.

Tiedeman, Robyne (1995). “Some Remarks on Antecedent Contained Deletion,” in Shosuke Haraguchi and Michio Funaki, *Minimalism and Linguistic Theory*, Hituzi Linguistics Workshope Series No. 4, Japan: Hituz Syobo.

Vergnaud (1974)

Wilder, Chris (1995). “Antecedent Containment and Ellipsis,” FAS Papers in Linguistics 4, 132-165.

Wilder, Chris (1997) “Phrasal Movement in LF: de re readings, VP Ellipsis and Binding,” *Proceeding of NELS 27*, 425-440.

Williams, Edwin (1974). *Rule Ordering in Syntax*, , Doctoral dissertation, MIT, Cambridge, MA: MITWPL.

Williams, Edwin (1977) “Discourse and logical form,” *Linguistic Inquiry* 8.101-139.

William Edwin (1995) “Ellipsis,” ms. Princeton Univeristy.

Wyngaerd, Vanden G. and C.J.W. Zwart. 1991. Reconstruction and Vehicle Change. In Linguistics in the Netherlands 1991, F. Drijkoningen and A. van Kemenade, eds. John Benjamins, Amsterdam, 151-160.

Danny Fox
Harvard University
Society of Fellows
78 Mt. Auburn st.
Cambridge, MA. 02138