Movement and Control

Norbert Hornstein

Since the earliest days of generative grammar, control has been distinguished from raising: the latter the product of movement operations, the former the result of construal processes relating a PRO to an antecedent. This article argues that obligatory control structures are also formed by movement. Minimalism makes this approach viable by removing D-Structure as a grammatical level. Implementing the suggestion, however, requires eliminating the last vestiges of D-Structure still extant in Chomsky's (1995) version of the Minimalist Program. In particular, it requires dispensing with the θ-Criterion and adopting the view that θ-roles are featurelike in being able to license movement.

Keywords: obligatory control, nonobligatory control, Minimal Link Condition, Greed

1 Introduction

This article is an exercise in grammatical downsizing. Since the earliest days of generative grammar (Rosenbaum 1967), control and raising constructions have been treated differently, with different rules and/or formatives involved in the two structures. In the beginning there was Equi-NP Deletion. Equi, a deletion process, contrasted with Subject-to-Subject Raising, a movement process. Subsequently, in most versions of the Extended Standard Theory, control was relegated to binding theory—the binding of an abstract expression PRO—whereas raising remained an instance of movement. This dual-track approach persisted into the Government-Binding (GB) era.

In GB, control sentences like (1a) have structures like (1b). These contrast with raising sentences, (2a), and their phrase markers, (2b). In particular, the relation between *John* and the embedded subject position in (1a) is mediated through the binding of a grammatically distinctive lexical formative in control configurations, namely, PRO. In raising structures like (2a) the relation between the matrix and embedded subjects is a by-product of movement and results in an A-chain in which the head, the antecedent, binds the tail, its trace.

(1) a. John expects to win.
    b. John_i expects [PRO_i to win]

(2) a. John seemed to win.
    b. John_i seemed [t_i to win]

This article went through several incarnations and has, I believe, risen to a higher karmic plane owing to the kind intervention of others. I especially thank Joseph Aoun, David Lightfoot, Jairo Nunes, and Juan Uriagereka. Special thanks to Rozz Thornton, who started me thinking along the lines outlined here. This work was supported by NSF grant SBR9601559.

Linguistic Inquiry, Volume 30, Number 1, Winter 1999
69–96

The differences do not stop here. The distribution of PROs in GB is attributed to binding theory—the PRO Theorem, to be precise. The distribution of NP-traces, in contrast, is the province of the Empty Category Principle (ECP). Traces must be properly governed. PROs, on the other hand, cannot be governed at all. PROs head their own chains; traces, by definition, cannot. PROs are base-generated; traces are produced through movement. Thus, in most every respect, GB fundamentally distinguishes NP-traces from PROs. Their one commonality within GB is that both are Caseless and phonetically null.

To date, standard work in the Minimalist Program has left matters pretty much in this GB state.[1] There are good reasons for this. Empirically, the distinction reflects the fact that the antecedent of PRO in cases like (1a) bears two θ-roles whereas the subject in (2a) has but one. This semantic difference is theoretically ensconced in the different *kinds* of binding assumed to hold in control versus raising. The theoretical basis for the distinction in GB technically rests on distinguishing a level of D-Structure. D-Structure is the sole locus of lexical insertion, an operation that precedes all other transformations. Lexical insertion is subject to θ-requirements. In particular, D-Structure is defined as the phrase marker that purely represents GF-θ, the level at which all and only thematic positions of the sentence are occupied by lexical material. Subsequent transformations move the lexical expressions located in θ-positions to non-θ-positions. These movements are further restricted by the θ-Criterion so that going from one θ-position to another is strictly forbidden.

This GB package of assumptions (the combination of D-Structure and the θ-Criterion) forces a distinction between PRO and trace, and thereby between binding and control. Two suppositions are central, and both are retained in the Minimalist Program: first, the θ-Criterion (the assumption that (A-)chains are constrained to possess but a single θ-position; i.e., movement from one θ-position to another is strictly forbidden); second, the priority of θ-marking over movement (i.e., the requirement that θ-positions coincide with the foot of a chain).

The first requirement prevents movement to θ-positions in the course of a derivation, just as it did in GB theories. The second retains a central feature of D-Structure. Chomsky (1995) operationalizes the thematic restriction on lexical insertion by restricting θ-assignment to the merger of trivial chains. This recapitulates within the Minimalist Program the assumption that D-Structure is the locus of pure GF-θ. Thus, in Chomsky's (1995) version of the Minimalist Program, a D/NP can legitimately enter a derivation only through the thematic door; that is, nominal expressions all enter the derivation via Merge. Given the provision that only trivial chains can be θ-marked, an NP so merged must merge to a θ-position on pain of never receiving a θ-role.

[1] This is not quite accurate. Recently control has become a hot area of research. My proposal shares with O'Neil's (1995) the intuition that control should be reduced to movement. Though the details of the two approaches differ, they are conceptually very similar. There are two other approaches to control set within minimalist assumptions. Martin (1996) develops a theory exploiting the notion of null Case proposed by Chomsky and Lasnik (1993) to account for the distribution of PRO. Manzini and Roussou (1997) develop a theory of control in terms of feature movement of heads at LF. The wealth of approaches is to be welcomed given the awkward position that the control module has in the Minimalist Program. In what follows I lay out a view of control that differs from those noted here. For reasons of space, I keep comparisons to a minimum.

Chomsky (1995) further assumes that all subsequent movement is restricted to nonthematic targets. This is technically executed by assuming (3).

(3) a. θ-roles are not features.
b. Movement must be greedy.

As θ-roles are not ''checkable'' features, movement to θ-positions cannot be greedy and so is prohibited. In short, the Minimalist Program retains the θ-Criterion.

All of this suggests that the minimalist abandonment of D-Structure as a level (Chomsky 1993) is less radical than often perceived. Chomsky's argument does not lead to a general repudiation of the core characteristics of D-Structure. Rather, D-Structure's earlier properties are packed into restrictions on the computational operations. In fact, the only feature of D-Structure that the Minimalist Program forswears is the principle that *all* lexical insertion precedes the application of *all* other transformations; in other words, the rule Satisfy has been dropped (Chomsky 1993). The other features of D-Structure have been retained.

This article submits these other assumptions to minimalist scrutiny. How well motivated are they? Why assume that chains are biuniquely related to θ-roles? What goes wrong if movement takes place from one θ-position to another? Why distinguish trace from PRO? As is generally the case with minimalist meditations, I assume that the burden of proof is on those who wish to promote these assumptions and invoke these distinctions. What is not at issue is that control and raising sentences manifest different properties. The minimalist question is whether these differences require the technical apparatus standardly invoked to distinguish them.

In the particular case of control, methodological skepticism is fully warranted. The distinction between raising and control multiplies the inventory of empty categories. Furthermore, the distinction massively complicates the grammar. PRO brings with it two big theoretical complications: (a) a control module whose job it is to specify how PRO is interpreted and (b) theoretical modifications to account for PRO's distribution. In GB (b) is handled by the binding theory. PRO is analyzed as a pronominal anaphor. The contradictory requirements that standard versions of the binding theory place on pronouns and anaphors within governing domains force such expressions to be ungoverned. Hence, PROs can appear only in ungoverned positions (see Chomsky 1986).

This GB approach to (b) has several conceptual and empirical problems (see Bouchard 1984, Chomsky and Lasnik 1993). Furthermore, the Minimalist Program cannot adopt the PRO Theorem, since it relies on the notion of government, which is not an acceptable minimalist primitive. Consequently, governing categories and domains cannot be defined or theoretically exploited.

Chomsky and Lasnik (1993) propose that the distribution of PRO is regulated by Case theory.[2] Theirs is the standard minimalist account for the distribution of PRO.[3] In particular, they propose that PRO has ''null'' Case. This is a Case special to PRO in the sense that only PRO bears it and Is that assign/check it license no other sorts of Case. It is fair to say that null Case accounts for the distribution of PRO largely by stipulation.

[2] This was first suggested by Bouchard (1984). Chomsky and Lasnik adopt Bouchard's basic proposal, though their particular version differs in certain details.

[3] This theory has been elaborated and expanded in Martin 1996.

The theory of the control module does not fare much better. What principles determine the antecedents of PRO, and whether or not all instances of control are actually the same, is quite controversial. It seems safe to say that control theory has not been one of the bright stars in the GB firmament.

In sum, neither part of the control conglomerate has been uncontroversial even within GB. Given a minimalist sensibility, its technical complexities are ripe for reevaluation.

The article is organized as follows. Section 2 reviews why we need a theory of PRO and control. In particular, sections 2 and 3 review the distribution and interpretive requirements of PRO in obligatory control (OC) and nonobligatory control (NOC) configurations. Sections 4 and 5 argue that the general properties of OC structures can be reduced to movement if we abandon the residues of D-Structure still extant within the Minimalist Program and abandon the θ-Criterion-based prohibition against moving into θ-positions. Section 6 briefly addresses how NOC and OC structures are selected. Section 7 concludes the exercise.

2 The Issues

Section 1 has tersely outlined the twin problems that theories of control must deal with: the distribution and interpretation of PRO. The GB solutions to these problems often pull in opposite directions. This tension is resolved in minimalist approaches. Consider the details.

The GB PRO Theorem treats PRO simultaneously as an anaphor subject to Principle A of the binding theory and a pronoun subject to Principle B. In the context of the binding theory, this dual status of PRO implies that it is ungoverned. Were it governed, it would possess a governing category and would have to meet both Principles A and B. As being both bound and free *within a given domain* is impossible, it must be that pronominal anaphors (i.e., PROs) evade this contradiction by not having governing categories. An XP with no governor (i.e., a governing X^0) is without a governing category. Thus, if PROs do not have governors (i.e., exclusively appear in ungoverned positions), they can meet their binding requirements. If one further assumes that the [Spec, IP] position of infinitives and gerunds is ungoverned, the observed distribution of PRO is explained.

This reasoning offers an additional bonus: it explains why PRO is phonetically null. In GB a D/NP is assigned Case by a Case-marking head that governs it. Since PROs cannot be governed, they cannot be Case-marked. If phonetically full D/NPs must be Case-marked, then PRO's lack of a governor precludes its having phonetic content. In sum, the PRO Theorem and Case theory combine to account both for PRO's distribution and for its phonetic status.

This GB account is less successful in dealing with PRO's interpretive characteristics. Williams (1980) argues that there are two types of control; obligatory control (OC) and nonobligatory control (NOC). This distinction, which has generally been accepted,[4] is based on several interpretive phenomena that distinguish OC from NOC constructions. Consider the following OC paradigm:[5]

[4] See, for example, Bouchard 1984, Koster 1984, Manzini 1983, Lebeaux 1985, Hornstein and Lightfoot 1987.

[5] (4a–e) are presented in Lebeaux 1985. (4f) is discussed in Higginbotham 1992. (4g) was first discussed in Fodor 1975.

(4) a. *It was expected PRO to shave himself.
b. *John thinks that it was expected PRO to shave himself.
c. *John's campaign expects PRO to shave himself.
d. John expects PRO to win and Bill does too. (= Bill win)
e. *$John_i$ told $Mary_j$ PRO_{i+j} to wash themselves/each other.
f. The unfortunate expects PRO to get a medal.
g. Only Churchill remembers PRO giving the BST speech.

(4a) shows that OC PRO must have an antecedent. (4b) indicates that this antecedent must be local, and (4c) that it must c-command the PRO. (4d) indicates that OC PRO only permits a sloppy interpretation under ellipsis. (4e) shows that OC PRO cannot have split antecedents. PRO in (4f) has only the *de se* interpretation, in that the unfortunate believes *of himself or herself* that he or she will be a medal recipient. In theory, (4g) could have two paraphrases, (5a) and (5b). In fact, however, only paraphrase (5a)—on which only Churchill could have this memory, for Churchill was the sole person to give the speech—is available. This contrast follows on the assumption that OC PRO must have a c-commanding antecedent. This requires *only Churchill* to be the binder.

(5) a. Only Churchill remembers himself giving the BST speech.
b. Only Churchill remembers that he gave the BST speech.

These properties of OC PRO are not shared by PRO in NOC environments.[6]

(6) a. It was believed that PRO shaving was important.
b. $John_i$ thinks that it is believed that PRO_i shaving himself is important.
c. Clinton's$_i$ campaign believes that PRO_i keeping his sex life under control is necessary for electoral success.
d. John thinks that PRO getting his resume in order is crucial and Bill does too.
e. $John_i$ told $Mary_j$ [that [[PRO_{i+j} washing themselves/each other] would be fun]].
f. The unfortunate believes that PRO getting a medal would be boring.
g. Only Churchill remembers that PRO giving the BST speech was momentous.

(6a) indicates that NOC PRO does not require an antecedent. (6b) demonstrates that if it does have an antecedent, the antecedent need not be local. (6c) shows that the antecedent need not c-command the NOC PRO. (6d) contrasts with (4d) in permitting a strict reading of the elided VP (i.e., the reading on which it is John's resume that is at issue). (6e) indicates that split antecedents are readily available in NOC contexts. (6f) can have a non–*de se* interpretation. (6g) is consistent with the notion that many people other than Churchill recall that the BST speech was momentous.

[6] It is even possible to ''control'' an NOC PRO across a sentence boundary. I believe that this was first noted by Emmon Bach, though I cannot recall where.

(i) $John_i$ even shaved for the interview. PRO_i making himself presentable is very important to the success of the project.

The OC cases in (4) and the NOC cases in (6) contrast in one further interesting way; it is possible to paraphrase the former by replacing PRO with reflexives, and the latter by replacing it with pronouns. (7) illustrates this with the counterparts of (4c) and (6c).

(7) a. *John's$_i$ campaign expects himself$_i$ to shave himself.
b. Clinton's$_i$ campaign believes that his$_i$ keeping his sex life under control is crucial for electoral success.

In short, the differences between OC and NOC structures duplicate, where applicable, the differences between structures with locally bound anaphors and pronouns (see (5a) and (5b) as well). This makes sense if PRO is actually ambiguous—an anaphoric expression in OC configurations and a pronominal in NOC structures—rather than simultaneously a pronoun and an anaphor, as the PRO Theorem requires. This, then, speaks against reducing the distribution of PRO to the binding theory by way of the PRO Theorem.

As noted in section 1, the PRO Theorem approach is theoretically unappealing from a minimalist perspective, relying as it does on government. Chomsky and Lasnik (1993) have further argued against this approach to PRO on empirical grounds. Consider the sentences in (8).

(8) a. We never expected [PRO to be found t].
b. *We never expected [PRO to appear to t [that Sally left]].

If movement is a last resort operation and PRO must be ungoverned, then the threat of being governed suffices to force PRO's movement in (8a). But if being governed suffices to license movement in (8a), why is it insufficient in (8b)? Chomsky and Lasnik argue that both examples are accounted for if one assumes that PRO has a Case that must be checked. Movement in (8a) is then a typical case of last resort movement under passive. (8b)'s unacceptability stems from a violation of Greed; since PP is a domain for Case checking, movement to [Spec, IP] is unnecessary and so prohibited. This essentially assimilates the unacceptability of (8b) to that of (9).

(9) *We never expect that Sally will appear to t that. . .

In (9) *Sally* raises to [Spec, IP] to check Case. However, it has moved from within PP, which is also a Case-checking domain. As Case has already been checked inside the PP, further movement is prohibited by Greed. Chomsky and Lasnik propose treating (8b) in analogous fashion. They assume that the embedded [Spec, IP] in (8b) is a Case position. PRO checks its Case here. Movement of PRO from within the PP to this [Spec, IP], therefore, violates Greed.[7]

Chomsky and Lasnik also trace the absence of lexical D/NPs in this [Spec, IP] to Case theory. These nonfinite specifier positions (Specs) can only check null Case, and null Case can only be carried by PRO. Thus, only PRO can appear in these positions.

The distribution of PRO, then, is relegated to Case theory, specifically to the Case properties of nonfinite ''control'' Is and null Case. The advantage of this solution from a minimalist point

[7] Adopting the assumptions in Chomsky 1995, the offending sentences would be ruled out because the Case of the embedded I^0 is not checked or because Case checking has resulted in feature clash within the PP. The movement does not violate Greed, because it checks the D-feature of the embedded I. However, I's Case feature cannot be checked by the raised PRO, either because it has been checked prior to raising within the PP or because the Case on PRO and the

of view is that it accounts for the distribution of PRO in terms of specifier-head feature checking.[8] This in turn permits the interpretive facts noted in (4) and (6) to be explained in terms of the ambiguity of PRO. Once the distributive properties of PRO are made the province of Case theory, the assumption that PROs are pronominal anaphors is no longer required. Rather, OC PRO is an anaphor and NOC PRO is a pronoun. At no time need PRO be both at once. The interpretive differences can now be seen to conform to whatever semantic properties differentiate anaphors from pronouns.

In sum, in contrast to GB's binding-theoretic approach, the Minimalist Program treats the distribution of PRO as a Case-theoretic phenomenon. This theoretical reapportionment, which is forced on both theory-internal and empirical grounds within the Minimalist Program, has as a side benefit the simplification of the theory of control. In particular, how PRO is interpreted can be related to whether PRO is anaphoric or pronominal. In OC cases it is anaphoric and displays the paradigm in (4). In NOC cases it is pronominal and displays the paradigm in (6). The differences between OC and NOC then follow on general grounds and so need not be stipulated as part of an independent control module.[9]

3 Some Problems

Despite the virtues reviewed above, the Case-theoretic account of the distribution of PRO suffers from certain inelegancies. The most glaring is that it essentially stipulates the distribution of PRO. Null Case is special in two ways. First, it is designed to fit only one expression: PRO. Lexical expressions do not bear null Case, nor do other phonetically null expressions such as *wh*-trace or NP-trace. Second, only nonfinite T^0s can check/assign it. In effect, the Case properties of PRO and nonfinite T are constructed to exactly fit the observed facts. Were the data otherwise, the theory would change accordingly. This comes close to restating the observations: PRO appears in the [Spec, IP] of a nonfinite I.

It may well be that this is the best that we can do theoretically. However, it is worth observing just how anemic an ''explanation'' this is. Particularly dubious is the stipulation that only PROs bear null Case. To date, only null Case distinguishes its bearer entirely in phonetic terms.

A further problem is that a null-Case-marked PRO fails to block contraction, unlike other Case-marked empty categories. Lightfoot (1976) observes that *wh*-traces block *wanna* contraction but NP-traces do not. Jaeggli (1980) accounts for this in terms of Case; Case-marked traces block the phonological phrasing of *to* with *want* to yield *wanna.*

(10) a. Who do you want [*wh*-trace to vanish]?
*Who do you wanna vanish?
b. John's going [NP-trace to leave].
John's gonna leave.

Case on PP do not match and the derivation terminates. At any rate, the unacceptability of the construction reduces to whatever blocks the derivation in (9).

[8] This is similar to how Chomsky (1993) treats the distribution of small pro: namely, by relating it to the feature structure of certain types of Is.

[9] The interpretive advantages of replacing the PRO Theorem with null Case are also noted by Boškovic (1995).

PRO patterns with NP-trace rather than *wh*-trace.

(11) I want [PRO to leave].
I wanna leave.

If PRO is a Case-marked expression, as the theory of null Case proposes, it should pattern like *wh*-trace and block contraction. That it behaves like NP-trace argues that it is more like non-Case-marked NP-trace than like Case-bearing *wh*-trace (see below).

A second problem with the Case-theoretic account is that it still requires a rather elaborate PRO module. This module functions to specify the antecedent, the ''controller,'' of the OC PRO. The module will of necessity be rather elaborate if it addresses the core empirical issues, of which there are two.

The first issue is this. In clauses that contain one nominal argument and a sentential argument, the subject controls the OC PRO.

(12) John_i hopes/expects/wants [PRO_i to leave].

When the matrix verb has two nominal arguments and a sentential PRO complement, the structure is typically one of object control rather than subject control.

(13) John_i persuaded Bill_j [$\text{PRO}_{*i/j}$ to leave].

In fact, subject control (e.g., in *promise* constructions) appears to be marked and emerges rather late in the acquisition process (see Chomsky 1969). The Minimal Distance Principle (MDP) captures the observations in (12) and (13) (Rosenbaum 1970). The MDP and the data it summarizes raise several questions: Why is OC the typical case in (13)? What sort of grammatical principle is the MDP? Is it a primitive generalization or does it follow from deeper principles? If the latter, what does it follow from? From a minimalist perspective, the MDP bears a striking resemblance to the Minimal Link Condition (MLC). Both prohibit structures like the *i*-indexing in (14) when there is a closer potential antecedent for the empty category (EC), be it PRO or NP-trace.

(14) $[\ldots \text{NP}_i \ldots \text{NP}_j \ldots \text{EC}_{*i/j} \ldots]$

(14) points to a possible redundancy between the MDP and the MLC and raises the obvious question: can they be collapsed into a single condition?

The second issue is this. Unlike complements, PRO-headed adjuncts do not permit object control.[10] Why is object control into adjuncts not permitted, and why is there a complement/adjunct asymmetry with regard to the controller of PRO in OC configurations?[11]

[10] This slightly overstates the case. There is object control into rationale clauses (as pointed out by a reviewer).

(i) They punished John for PRO driving drunk.

It is plausible that this adjunct is attached lower than those headed by *after, before, while,* and so on, thereby permitting control into them. For discussion of adjunct control, see section 5.

[11] The reason that is implicitly assumed is that objects, unlike subjects, cannot bind into adjuncts (see Chomsky 1995:272ff.). The empirical basis for this assumption is not very clear. Note that objects can license bound pronouns. As these are typically thought to be licensed under c-command, it appears that objects can in fact bind into adjuncts.

(i) Mary greeted every boy_i without his_i knowing it.

(15) John$_i$ saw Mary$_j$ without PRO$_{i/*j}$ leaving the room.

A third problem with the Case-theoretic account is that there appear to be cases of control where PRO is in a position other than [Spec, IP].[12]

(16) John$_i$ washed/dressed/shaved (PRO$_i$/himself$_i$).

The examples in (16) relate a thematic subject, *John*, to a thematic object. This object can be lexically realized or not. When it is not overt, the sentence is interpreted reflexively. These are the properties we would expect if the object were an OC PRO. Note that, as in (4a–g), the OC PRO alternates with a locally bound reflexive. Under standard assumptions, the object position cannot contain any empty category other than PRO given that its antecedent appears to bear a θ-role. However, cases like (16) are problematic since they suggest that PRO can appear in positions other than [Spec, IP]. If this is correct, then Chomsky and Lasnik's Case-theoretic account is incomplete, and the relevant issue is not the null-Case-marking/Case-checking capacities of non-finite T^0.

A fourth problem is that the account of control sketched in section 2 exploits the properties of anaphors in accounting for the basic facts about OC. The theory of anaphora, it appears, is also in need of some minimalist rethinking. The standard GB version in Chomsky 1986 exploits locality notions (e.g., governing category) of dubious standing in the Minimalist Program. Other approaches introduce redundancies into the grammar by crucially exploiting locality conditions characteristic of movement (e.g., the Chain Condition in Reinhart and Reuland 1993). Such redundancies are partially alleviated by requiring anaphors to move at LF. However, this raises other problems in a minimalist context, for (like all other forms of movement) this LF operation requires featural motivation. The problem is that it is unclear what features this movement would be checking. Anaphors appear to have Case, so movement to the antecedent would not be motivated as it is in existential constructions, for example. Another spur to movement could be the requirement to check ϕ-features. However, if these are interpretable (Chomsky 1995), then they need not be checked. Furthermore, if it is ϕ-feature checking that is at issue, why do reflexives not appear in finite-clause subject positions? What would block the (LF) movement in (17) if reflexives moved at LF to their antecedents to check ϕ-features?[13]

(17) John thinks that *heself/*himself is handsome.

In sum, the theory of anaphora raises problems for the Minimalist Program. It would therefore be good to be able to deal with the properties of OC without relying too heavily on any particular current account of anaphora.

Finally, given minimalist inclinations, the deepest question concerning PRO is whether such

[12] Lasnik (1995a) discusses these cases and attributes his observations to Alan Munn. Lasnik, like Munn, suggests that these cases be treated in terms of movement. The account in the text develops this suggestion for all cases of OC.

[13] Chomsky (1986:176) blocks this case in terms of the ECP: movement of the reflexive at LF violates this condition. The details of this proposal are not replicable in a minimalist framework as it appears to rely on lack of proper government to block the derivation. See Hornstein 1995 for discussion of the place of the ECP in the Minimalist Program.

a formative even exists. PRO is a theory-internal construct.[14] In GB, PRO is structurally analogous to NP-traces and *wh*-traces. All have the same shape, namely, [$_{NP}$ e]. The main difference between traces and PRO is the source of their indices: the former derive from movement, the latter are assigned via the control module. In the Minimalist Program, however, this machinery is all suspect. There is little reason to think that traces (qua distinctive grammatical constructs) exist at all. Traces are not grammatical formatives but the residues of the copy-and-deletion operations necessary to yield PF/LF pairs. As such, traces have no common structure in the Minimalist Program as they do in GB. They are simply copies of lexical material and so have no specific shapes whatsoever. Thus, they cannot be structurally analogous to PRO. This leaves the theoretical status of PRO up in the air. What kind of empty category is it? Why do grammars have it?

Section 1 has provided answers to these questions. PRO exists because of θ-theory. If chains could bear more than one θ-role and if θ-roles could be accreted in the course of a derivation, there would be little reason to distinguish PROs in OC configurations from NP-traces. As these restrictions on θ-assignment are not conceptually necessary, the theoretical basis for distinguishing PROs from NP-traces weakens. Put more bluntly, distinguishing trace from PRO requires *additional* assumptions about θ-assignment and chains. The burden of proof, therefore, resides with those who favor such assumptions. In section 4 I argue that forgoing these stipulations permits a more empirically and theoretically adequate account of OC. I propose that PRO, like NP-trace, is the residue of movement. Strictly speaking, then, there is no grammatical formative like PRO. Rather, PRO is simply a residue of movement—simply the product of copy-and-deletion operations that relate two θ-positions.[15]

4 An Alternative

I have argued that the null hypothesis is that OC PRO is identical to NP-trace; that is, it is simply the residue of movement. NOC PRO is to be identified with pro, the null pronominal found in various Romance and East Asian languages. This section is concerned with demonstrating the empirical virtues of these assumptions. The main focus is on OC PRO, since handling the OC data requires the most radical departures from standard GB and minimalist technicalia. For what follows, I adopt the following assumptions:

(18) a. θ-roles are features on verbs.
b. Greed is Enlightened Self-Interest.
c. A D/NP ''receives'' a θ-role by checking a θ-feature of a verbal/predicative phrase that it merges with.
d. There is no upper bound on the number of θ-roles a chain can have.
e. Sideward movement is permitted.

[14] See Nunes 1995 for a discussion of the status of traces as formatives in the Minimalist Program.

[15] It is interesting to observe that earlier theories of control that distinguished OC and NOC assume that OC PRO is governed (see, e.g., Manzini 1983, Hornstein and Lightfoot 1987). In many versions of the ECP, government by a head is required for all empty categories resulting from movement (see, e.g., Aoun et al. 1987, Rizzi 1990). The fact that OC PROs are the head-governed ones once again suggests that they, like traces in general, are the residues of movement.

(18a) treats θ-roles as morphological features.[16] This is required if movement to a θ-position is to conform to the principle of Greed. If OC is to be reduced to movement, then this assumption is conceptually required given other minimalist assumptions. (18b) interprets Greed as requiring at least one of the relata to check a feature (Lasnik 1995a). Thus, if A moves to merge with B, then at least one feature of either A or (the head of) B is checked. Treating θ-roles as features on the verb or predicate allows a D/NP to move to a θ-position and respect Greed by checking this feature.[17] Analyzing θ-roles thus permits us to ''mechanize'' θ-role assignment as in (18c): to receive a θ-role is just to check the relevant thematic feature of the predicate. One might think of this as ''transferring'' the verbal θ-feature to the nominal expression. In effect, checking conforms to Chomsky's (1995:226) vision of syntactic operations as the ''rearrangements of properties of the lexical items of which they are ultimately constituted''—that is, the features of the elements in the array. (18d) is logically required to analyze OC in terms of movement given that control involves the relation of at least two θ-positions. It is also the null hypothesis, I believe. The requirement that chains be restricted to a single θ-role needs substantial empirical justification. (18e) comes into play in the analysis of adjunct OC. I discuss it further in that context. What is important here is that c-command is not part of the *definition* of movement. Thus, the computational system does not prohibit the copying of an expression to a position that does not c-command the ''movement'' site.[18]

The assumptions in (18) suffice to accommodate OC in terms of movement given standard minimalist technology. Their empirical virtue is that they permit a radical simplification of the grammar of control and a derivation of the basic properties of OC structures. Consider the details.

First, consider the basic interpretive properties of OC structures. As noted in section 2, these structures require c-commanding local antecedents (see (4a–c)). This is what one expects if OC PROs are NP-traces. For illustration, let us look at (19).

(19) a. John hopes to leave.
b. [$_{IP}$ John [$_{VP}$ John [hopes [$_{IP}$ John to [$_{VP}$ John leave]]]]]

The derivation begins with *John* merging with *leave,* thereby checking the verb's θ-role. *John* then ''raises'' to the embedded [Spec, IP] to check the D-feature of the IP. This is *not* a Case-marking position, so the Case of *John* cannot be checked here. *John* raises again to [Spec, VP] of *hope* and checks the external θ-feature of the verb. By (18c), each time *John* checks a θ-feature of a predicate, it assumes that θ-role. Thus, *John* (or the chain it heads) has two θ-roles, the leaver role and the hoper role. *John* raises one last time to [Spec, IP] of the matrix, where it checks the

[16] This has already been proposed by Lasnik (1995a), Boškovic (1994), and Boškovic and Takahashi (1998).

[17] Chomsky (1995) suggests that it is odd to think of θ-roles as features. This is correct if one thinks of them as properties of D/NPs. There is no ''paradigm'' that groups nominals by their thematic status. However, verbs (and other predicates) are indeed grouped by adicity. In other words, verbs are categorized by their thematic status. This makes it quite natural to treat θ-roles as features of predicates as proposed here.

[18] See Nunes 1995 for elaborate discussion. This option does not prohibit making c-command a property of chains, however. Thus, sideward movement might be allowed, but the resulting copies might have to be in a c-command sequence to be interpretable objects. I am not advocating this here, however. The negative thesis suffices for the time being. See section 5 for discussion of sideward movement.

D-feature of the IP and nominative Case. Note that this is the only place where *John* checks Case. On the assumption that it was inserted into the derivation with nominative Case features, the derivation converges. In more conventional notation, the copy *John* in the embedded [Spec, IP] corresponds to PRO, and the copy in the matrix [Spec, IP] is the antecedent. The requirement that OC have a local c-commanding antecedent follows from the fact that PRO is an intermediate link in an A-chain. As such, it must have an antecedent. Furthermore, the antecedent must conform to general A-chain strictures and thus both c-command the traces in the A-chain (i.e., the PRO in [Spec, IP]) and be local to it, given conditions on movement like the MLC. In short, the first three properties of OC PRO follow straightforwardly (see (4a–c)).

Treating OC PRO as the residue of movement also derives the prohibition against split antecedents. Two (nonconjoined) expressions cannot both antecede OC PRO because they cannot have both moved from the same position. In other words, the ban against split antecedents in this case is equivalent to the ban against one and the same trace having two distinct antecedents. In the Minimalist Program this reduces to the fact that two distinct expressions cannot be merged into a single position.[19]

The required sloppy reading of OC PRO follows as well. Note that in raising constructions only a sloppy reading is available.

(20) Mary seems to be happy and Sally does too.

(20) must be understood to mean that it seems that Sally is happy. For the same reason, OC PRO must carry the sloppy reading since it too is an NP-trace.[20]

The movement analysis also accounts for the required *de se* interpretation of OC PRO. The movement underlying OC PRO ends up assigning two θ-roles to a single expression; for example, in (19a) *John* has two θ-roles. The semantic form of the predication in (19) is equivalent to (21), a predication that ascribes a reflexive property to the subject *John*.

(21) John λx [*x* hopes *x* leave]

Movement, then, semantically forms a compound monadic predicate by having one and the same expression saturate two argument positions. Salmon (1986) discusses these semantic issues at some length. Of importance here is his observation that relating the semantic value of an expression to two θ-positions via the formation of a reflexive predicate is semantically very different from relating two expressions in different θ-positions to each other via coreference. The former operation results in changing the semantic argument structure of the predicate; the latter leaves it intact. The former operation reflexivizes the predicate and thus forces a *de se* reading; the latter does

[19] This actually follows from the definition of Merge together with the analysis of movement as copy and deletion and of traces as copies. The fact that it does so is interesting, for in a representational theory that treats traces as real base-generated formatives (one version of GB), this prohibition against split antecedents must be stipulated.

[20] This does not say how sloppy readings are derived, whether interpretively (Sag 1976, Williams 1977) or via deletion operations at PF under some sort of parallelism requirement (Chomsky 1995). For current purposes, which of these proves to be correct is irrelevant. All that is required is that the ultimately correct approach should treat movement in an interpretively uniform fashion.

not. Treating OC as the reflex of movement, then, yields the correct interpretation for the structures—the one exemplified in (21).

Finally, the observed reading in (4g) (repeated here) follows as well.

(22) Only Churchill remembers giving the BST speech.

The reading on which someone other than Churchill could recall this event requires the paraphrase in (23).

(23) Only Churchill remembers Churchill giving the BST speech.

This cannot underlie the structure of (22). The PRO here is of the OC variety. This means that *only Churchill* has raised from the embedded position and has the reflexive property noted in (24). This is semantically equivalent to the reading on which Churchill alone has the required memory.

(24) only Churchill λx [x remembers x giving the BST speech]

In sum, the six basic properties of OC reviewed in section 2 follow directly from assuming that OC PRO is identical to NP-trace, the residue of movement. In addition, these properties are derived without the problems reviewed in section 3. Once again, consider the details.

The distribution of OC PRO does not require the services of null Case. This Case, specially designed for PRO by Chomsky and Lasnik (1993), is unnecessary if OC PRO is an NP-trace. In fact, the existence of null Case in [Spec, IP] of control infinitives is *incompatible* with the movement analysis, since it would prevent raising out of the embedded [Spec, IP].

Abandoning null Case in this context does not lead to any empirical difficulties. Recall that null Case has been postulated to replace the assumption that PRO must be ungoverned. Its principal empirical effect is to block the derivation of (25) and license PRO only in [Spec, IP] of nonfinite clauses.

(25) *We never expected [PRO_i to appear to t_i that . . .

The proposed account rules (25) out on the same basis as an account that postulates null Case. On the latter view, PRO cannot move to [Spec, IP] of the embedded clause because it would be moving from one Case-marking position (inside PP) to another ([Spec, IP]). This either violates Greed or causes a feature mismatch. In either case the derivation fails to converge. However, if [Spec, IP] is an *intermediate* NP-trace, as it would be on the proposed account, then the same reasoning prohibits movement through this position. In effect, a PRO in (25) should be no better than an NP-trace in (26).

(26) *We_i were expected [t_i to appear to t_i that . . .

Furthermore, on the proposed account we expect to find OC PRO in positions from which movement is licit. This should roughly coincide with non-Case-marked positions, such as [Spec, IP] of nonfinite clauses. Note that this is compatible with treating inherently reflexive verbs like *wash, dress,* and *shave* as simply not Case-marking their objects—in effect, as allowing derivations like (27) to be licit.

(27) a. Mary washed.
b. [$_{IP}$ Mary [past [$_{VP}$ Mary [wash Mary]]]]

Case is checked in [Spec, IP]. *Mary* receives two θ-roles since it checks both the internal and external θ-role of *wash*.[21]

In effect, then, by assuming that PRO is identical to an NP-trace, an intermediate NP-trace to be exact, we derive its distribution without having to assume null Case. This account has two further benefits. First, it allows us to treat *wanna* contraction over PRO and NP-trace as one and the same phenomenon (see (10) and (11)).[22] Second, the null phonetic status of PRO is explained in whatever way we explain the null phonetic status of NP-trace. One natural assumption is that Case is required for phonetic ''visibility.'' Both NP-trace and PRO will therefore fail to meet the requirements for having phonetic content.[23]

The movement approach to OC PRO also accommodates the classical data used to distinguish raising from control. It was argued, for example, that idiom chunks and expletives could raise but not control.

(28) a. The shit seems [t to have hit the fan].
b. There seems [t to be a man in the garden].

(29) a. *The shit expects [PRO to hit the fan].
b. *There expects [PRO to be a man in the garden].

The distinction between these cases is preserved in the present account even if PRO in (29) is just an NP-trace. The basis for the distinction is that in (28) *the shit* and *there* bear the external θ-role of *expect*. If this θ-role is not checked, then, I assume, the derivation fails to converge since there is an unchecked θ-feature at LF.[24] However, the only nominals that can check the

[21] The question remains, why this operation is not generally valid. If we assume that the difference between *wash* and *see* is that the latter must assign accusative Case whereas the former can suppress it, then we expect (i) to violate Case theory.

(i) John saw PRO.

The underlying structure would leave either the accusative Case of *saw* or the nominative Case of IP unchecked.

Note that if this is correct, then *expect* is like *wash* in optionally suppressing its accusative Case, and *believe* is like *see*. This permits the contrast in (ii).

(ii) a. John expects PRO to be leaving.
b. *John believes to be leaving.

[22] We similarly explain why it is governed PRO that Manzini (1983) and others took to be the OC PRO. The ECP is an empirically useful diagnostic of movement, whatever its theoretical shortcomings in minimalist terms. The reason that OC PRO is governed is that it is the residue of movement; that is, it meets the same descriptive requirements as other traces.

[23] This is essentially the suggestion made by Jaeggli (1980). Whether it is correct is not relevant here. All that is required is that PRO's phonetic status be accounted for in the same way as NP-trace's. For another approach, see Nunes 1995.

[24] If one is reluctant to require that θ-features be checked, one might recast the θ-Criterion to require that every θ-role be ''expressed'' by being attached to a DP. This would then make (29a–b) θ-Criterion violations. Note that the analysis proposed above requires dropping the assumption that a DP can bear only one θ-role. It does not require dropping the assumption that all θ-roles must be assigned.

relevant θ-roles, *there* and *the shit,* are not expressions that can bear θ-roles because of their inherent idiomatic or expletive semantics. As a result, we retain a difference between raising and control structures in cases such as these but attribute it not to an inability to control PRO but to an inability to support a θ-role that must be discharged for grammaticality to ensue.

This section has demonstrated that OC structures can be treated in terms of movement and that there is considerable empirical payoff in doing so. In particular, we can dispense with null Case, and we can derive the six basic properties of OC exemplified in (4). The next section turns to perhaps the biggest advantage. It appears that treating OC PRO as the residue of movement comes very close to allowing us to eliminate the PRO module entirely.

5 The Minimal Distance Principle Reduced to the Minimal Link Condition

The PRO module has two primary functions. First, it designates the controller in an OC structure. Second, it determines how a controlled PRO is to be interpreted in a given configuration (e.g., Does it permit split antecedents? Is it obligatorily *de se*?). The latter function of the PRO module is no longer required. The various interpretive options of OC and NOC PRO follow from whether the PRO in question is a null pronominal—pro—or a residue of A-movement—an NP-trace.[25] Still to be explained is how the controller in OC cases is determined. Note that this is not an issue for NOC configurations, since in these cases no antecedent is required. This is the topic of this section.

The chief descriptive principle regulating this part of the PRO module is the Minimal Distance Principle (MDP).[26] Its effect is to designate NP_i the controller in the configurations in (30).

(30) a. NP_i [V [PRO. . .]]
b. NP [V NP_i [PRO. . .]]

Thus, the MDP picks the closest c-commanding potential antecedent as controller; that is, α is the controller of PRO iff α c-commands PRO, and for all β different from α that c-command PRO, β c-commands α. The MDP picks the subject to be the controller in (30a) and the object in (30b). If the MDP is treated as a markedness condition, then verbs like *persuade* become the unmarked case and verbs like *promise* are highly marked. This state of affairs closely coincides with the observed data and I assume its accuracy here.

The MDP makes perfect sense from the perspective of a movement approach to OC. We have already seen the derivation of a structure like (30a) in (19b). Now consider the derivation of an object control sentence.

(31) a. John persuaded Harry to leave.
b. [$_{IP_2}$ John [$_{I^0}$ past [$_{VP_3}$ John v + persuaded [$_{VP_2}$ Harry persuaded [$_{IP_1}$ Harry [to [$_{VP_1}$ Harry leave]]]]]]]

[25] There remains the issue of when pro-headed propositions are permitted. I return to this in section 6.
[26] This was first formulated in Rosenbaum 1967.

The array consists of the following set of expressions: {John, Harry, persuaded, to, leave, past, v, other assorted functional categories}. The derivation starts by selecting *leave* and *Harry* and merging them. This allows *Harry* to check the θ-feature of *leave* and assume the internal argument role—(18a,c). *To* then merges with the VP headed by *leave,* and *Harry* moves to [Spec, IP_1] to check the D-feature of the embedded clause (the Extended Projection Principle).

Note that this move violates Procrastinate, since *John* could have been inserted here. However, if *John* had been inserted, the derivation would not have been able to converge. I return to the details after limning the rest of the derivation.

IP_1 then merges with *persuaded,* checking the propositional θ-role of the verb. *Harry* then raises and merges with VP to form [Spec, VP_2]. This too is a θ-position of *persuade,* and this move provides *Harry* with a second θ-role. Once again Procrastinate is violated, since *John* could have been inserted. Had it been, however, the derivation would have failed to converge, so the insertion is blocked (see below). The next step is to raise *persuaded* to merge with *v*. Then *John* is taken from the array and merged with the *v* + *persuaded* projection, forming [Spec, VP_3]. This is a θ-position, and *John* checks the external θ-role. Past tense features then merge with VP_3. *John* raises and forms [Spec, IP_2]. Here the D-features and nominative Case features of T are checked, as are the Case features of *John.* At LF *Harry* raises and forms an outer [Spec, VP] (or alternatively merges with Agr_O and forms [Spec, Agr_O]), where it checks its Case features and those of *v* + *persuade.* All features that must be checked are checked, and the derivation converges.

Observe that [Spec, IP_1], the so-called position of PRO, is occupied by an intermediate copy of *Harry.* This expression has been inserted with Case features. Assume that it has accusative Case; otherwise, the derivation does not converge. Each move in the derivation is licit with respect to Greed because some feature is checked at every step. The two violations of Procrastinate must still be accounted for, however. Let's turn to them now.

Harry is inserted into the derivation with some Case features.[27] If these features are accusative, then there is no way to check the features on *John.* There are two possibilities. If *John* has accusative features, then either its features or those on *Harry* cannot be checked, since only one accusative head is available: *v* + *persuaded.* If the features on *Harry* are nominative, then the Minimal Link Condition (MLC) will prevent movement of *Harry* across *John* if *Harry* is inserted into [Spec, IP_1] and raised again. The full LF phrase marker given this derivation is (32).

(32) [$_{IP_2}$ Harry [$_{I^0}$ past [John [$_{VP_3}$ John v+persuaded [$_{VP_2}$ John persuaded [$_{IP_1}$ John [to [$_{VP_1}$ Harry leave]]]]]]]]

[27] The standard assumption, adopted here, is that nominal expressions cannot be doubly Case-marked. This is also the minimal assumption given standard minimalist reasoning: universally, nominals require Case. This is minimally satisfied with one set of Case features. As no more are required, no more are permitted, at least in the unmarked case.

To check nominative Case requires moving *Harry* to I^0. This traverses several copies of *John,* all of which are closer. This violates the MLC and so is illicit (see below).

Consider the second option. *John* is inserted with nominative Case, *Harry* with accusative. *John* is merged into [Spec, IP_1]. It then raises through the two θ-positions of *persuade* and *v* up to [Spec, IP_2], where it checks its nominative Case and that of T, as well as the D-feature of IP_2. The accusative features on *Harry* need to be checked. This could be done by moving to the outer [Spec, VP_3]. The relevant LF phrase marker is (33).

(33) [$_{IP_2}$ John [$_{I^0}$ past [Harry [$_{VP_3}$ John v+persuaded [$_{VP_2}$ John persuaded [$_{IP_1}$ John [to [$_{VP_1}$ Harry leave]]]]]]]]

The derivation in (33) must be illicit if the one in (31b) is well formed as required by the present analysis. The derivation in (33) violates the MLC on the assumption that *John* in [Spec, IP_1], [Spec, VP_2], or [Spec, VP_3] prevents the movement of *Harry* in [Spec, VP_1] to the outer [Spec, VP_3] (or to Agr_O to form [Spec, Agr_O]) to check accusative Case at LF. This, in turn, requires that copies formed by movement be relevant for the MLC. *Harry* cannot check its Case because it is farther from the relevant Case-checking position than are *copies* of *John.* If we make the assumption that such copies are visible to the computational system (and thereby block movement across them via the MLC), then the derivation in (33) does not converge. This, in turn, permits Procrastinate to be violated in the derivation of (31b).

Interestingly, the assumption that copies are relevant for the MLC is required independently once the assumption that only trivial chains can be θ-marked—(18d)—is dropped. Given the assumption (18d) that there is no upper bound on the number of θ-roles an expression can have—a conceptually necessary assumption if OC is to be reduced to movement—(34) provides independent motivation for making the MLC sensitive to copies. Chomsky (1995:345) asks why (34b) does not exclude (34c).

(34) a. I expected someone to be in the room.
b. [$_{IP_1}$ I expected [$_{IP_2}$ I to be [someone in the room]]]
c. [$_{IP_1}$ I expected [$_{IP_2}$ someone to be [someone in the room]]]

The derivation of (34b) proceeds as follows. The small clause *someone in the room* is constructed via successive mergers. The result is merged with *be* and then with *to. I* is merged in the [Spec, IP_2] position, discharging the D-feature on the embedded I. The result is merged with *expected.* Chomsky argues that this structure is illicit because *I* cannot receive a θ-role given either the assumption that only trivial chains can be θ-marked or the assumption that θ-features do not count for Greed so that movement via the [Spec, VP] of *expect* is illicit. If convergence requires nominals to have θ-roles, then (34b) does not converge. This then licenses the derivation in (34c) in which Procrastinate is violated. In other words, instead of merging *I* to [Spec, IP_2], *someone* can be raised, violating Procrastinate. This is how (34a) is licitly derived.

The assumptions in (18) preclude adopting this analysis. Given (18b,d), there is nothing that prohibits moving *I* to [Spec, VP] of *expect* to receive a θ-role. Thus, (34b)'s failure to converge

cannot be due to the requirement that nominals have θ-roles. However, the derivation can be excluded in the same way that (33) is. *Someone* needs to check its Case features. The only available Case position is the outer Spec of *expected* (or the Agr_O above it). Movement to this outer Spec is blocked by *I* in [Spec, IP_2] if copies count for the MLC.[28] In short, the same reasoning required for (33) extends to cover this case as well.

(34) is of particular interest, for it indicates that treating copies as the computational equals of originals from the array is virtually unavoidable if OC is reduced to movement in the context of the Minimalist Program. (34b) must be prohibited from converging. θ-theory is unavailable once (18b,d) eliminate the last vestiges of D-Structure. In particular, there is nothing amiss with the derivation in (34b) if *someone* can check Case. Thus, it must be that it cannot. This is accomplished if the *I* in [Spec, IP_2] triggers the MLC.[29]

The above has been in the service of a single conclusion: that object control can be derived via movement given a general minimalist setting amended by (18). There is a further conclusion. Subject control in *persuade* clauses is ungrammatical. To derive a structure of subject control involves violating the MLC. Consider the derivation of a subject control structure like (35a).

[28] Note that, strictly speaking, the *I* in the embedded [Spec, IP] is not a copy but the original element selected from the array. The higher *I*s are the copies if movement is copying plus deletion as Chomsky (1995) assumes. The point in the text is that the grammar does not (and should not) distinguish copies from originals in any relevant sense. This is contrary to Chomsky's proposal that the foot of a chain differs from the head in not being visible to the computational system. Chomsky's proposal amounts to encoding in minimalist terms a distinction between expressions and their traces. In effect, it implicitly postulates the existence of traces as grammatical formatives. As usual, the postulation of abstract entities must be empirically justified. The null position is that NP-traces, qua distinctive grammatical objects, do not exist. This is what the copy theory presupposes and what I assume here. For further critical discussion of this assumption, see Nunes 1995.

Treating all copies as grammatically equal raises the question of whether chains are ''real'' objects, that is, have distinctive properties of their own. When introduced in Chomsky 1981, chains were a notational shorthand used for summarizing the properties of local movement. Rizzi (1986) was the first to argue that chains had an independent grammatical existence. In the context of the Minimalist Program, it is not at all clear that chains should be treated as independent entities. For example, their existence appears to contradict ''inclusiveness'' (Chomsky 1995:228), which bars the addition of ''new objects'' in the course of the computation from the numeration to LF. Chains are not lexical objects. Therefore, inclusiveness should bar their presence at LF. This is not to deny that movement exists. The existence of ''displacement'' operations in the grammar is undeniable. However, this does not imply that chains exist, with well-formedness conditions of their own. For further discussion of these issues, see Hornstein 1998b.

[29] The reasoning in (34) is the sole argument given in Chomsky 1995 for assuming that NPs must have θ-roles for a derivation to converge. The reasoning in terms of Case removes this argument.

There is another problem, however, that leads to the same conclusion. Consider (i) (with the structure in (ii)), brought to my attention by Juan Carlos Castillo (see also Martin 1996:26, (21)).

(i) *John expects to seem that he is smart.

(ii) $[_{IP_1}$ John $[_{VP}$ John expects $[_{IP_2}$ John to seem $[_{CP}$ that he. . .]]]]

According to the derivation in (ii), *John* is first merged into [Spec, IP_1]. It raises to [Spec, VP], checking the external θ-role of *expect,* and then raises to [Spec, IP_1] to check Case and D-features. This derivation is ruled out if movement cannot take place to a θ-position or if only trivial chains can receive θ-roles. However, these assumptions are rejected here, and so another way must be found to prevent the derivation in (ii). Two solutions are possible.

The first option is to deny that *John* can be inserted into [Spec, IP_2] since it fails to check some relevant feature. One possibility is that the I^0 of a control predicate differs from the I^0 in a raising construction. This is proposed by Martin (1996:chap. 2), following earlier work by Stowell (1982) on the tense of infinitives. The proposal is that control and raising infinitives have a [T] feature. The two types are distinguished in that control infinitives have [+T] I and raising infinitives have [−T] I. Say that this is correct. We now make a further assumption concerning [+T] infinitives: namely, that the [+T] control infinitive I (in contrast to the [−T] raising infinitive I) has a feature that can only be checked by

(35) a. $John_i$ persuaded Harry [PRO_i to leave].
b. [$_{IP_2}$ John [$_{I^0}$ past [$_{VP_3}$ John v + persuaded [$_{VP_2}$ Harry persuaded [$_{IP_1}$ John [to [$_{VP_1}$ John leave]]]]]]]

The relevant structure is (35b). If *John* is the controller, it must have been merged with *leave* and raised to [Spec, IP_1], the locus of ''PRO.'' *Harry* has the object-of-*persuade* θ-role, as indicated by its merger in [Spec, VP_2]. Now, the external θ-feature of *v* + *persuaded* must be checked. In (35b) *John* is raised to [Spec, VP_3]. Note that it crosses *Harry* in [Spec, VP_2]. This violates the MLC and is prohibited. As noted earlier, this is consistent with the traditional observation (see Chomsky 1969) that subject control verbs like *promise* are highly marked.

To conclude: I have argued that a derivational approach to OC can account for the prevalence of object control with *persuade* verbs. This is what is expected given minimalist technology supplemented with the assumptions in (18). In effect, OC structures conform to the traditional

---

an NP that has a θ-feature (i.e., is θ-marked). For concreteness, assume that the D-feature associated with [+T] infinitives can only be checked by a DP with a θ-feature. Recall that the account I am proposing treats θ-roles as features that an expression receives by checking the θ-features of a predicate under merger. In effect, the θ-features of the verb are transferred to the DP that merges with it. If they are indeed features, then we would expect them to enter into typical checking relations. The proposal is that this is what happens in control IPs; that is, only DPs having θ-features can check the I of a [+T] OC infinitive. Note that this will prevent the derivation in (ii), for *John* merged into [Spec, IP_2] has not yet been θ-marked. It receives a θ-feature only by moving to [Spec, VP]. Thus, it cannot check the postulated θ-sensitive feature of the [+T] infinitive I. Note that this same assumption will suffice to block control by expletive arguments like *it* in sentences like (iii).

(iii) *It was hoped to be believed that Fran left.

(iii) cannot be interpreted as parallel to (iv).

(iv) It was hoped that it was believed that Fran left.

The question is, why not? The relevant structure is (v).

(v) [$_{IP_2}$ it was hoped [$_{IP_1}$ it to be believed [that. . .

This is illicit if we assume that *it*—not endowed with a θ-role (or endowed with the ''wrong kind'' of pseudoargument θ-role in the sense of Chomsky 1986)—cannot check the postulated feature of the embedded [+T] infinitive that *hope* selects. Note that if PRO were a formative, it is not clear how (v) would be ruled out except by stipulating that it cannot be bound by a non-θ-marked antecedent in these sorts of constructions. This, however, runs against the observation that such control is possible in adjunct structures (see Chomsky 1986, where (via) is discussed).

(vi) a. It_i always rains [before PRO_i snowing].
b. It_i seemed that Clinton won reelection without PRO_i appearing that he had won a majority.

In short, postulating a feature on [+T] infinitives that must be checked by θ-marked DPs suffices to handle (i).

A second option for dealing with (i) is to recognize that when predicates like *seem* take finite complements, they require *it* subjects. If, as Chomsky (1986) suggests, *it* carries a weak kind of θ-role, then one can treat this requirement as a selection fact about predicates with finite complements: they select thematically marked subjects like *it.* If so, then (i) is ruled out because it violates a selection restriction. Note that the intimate relation between *it* and finite complements has long been observed. The proposal here is to exploit that fact to block the derivation in (i). A similar sort of account would block the derivation of (vii), which Chomsky (1995) treats as a selection restriction violation.

(vii) *I expect [$_{IP_1}$ John to seem [that t left]].

This raising should be licensed by the requirement to check the D-feature in IP_1. It is blocked if the finite complement requires that *seem* have an *it* subject.

For current purposes, either approach suffices. I assume that one of them can be more fully worked out to accommodate the problematic (i); thus, (i) does not threaten the basic assumptions required for a movement theory of OC.

MDP just in case their derivational histories respect the MLC. The control module and MDP are superfluous in cases like these.

Consider now the absence of object control in structures like (36), illustrated in (37).

(36) NP_i V NP_j [$_{Adjunct}$ $PRO_{i/*j}$ . . .]

(37) $John_i$ heard $Mary_j$ [without/before/after [$PRO_{i/*j}$ entering the room]].

The account for these structures can also be reduced to the MDP if it is assumed that objects do not c-command adjuncts. In minimalist terms, this requires assuming that objects fail to c-command adjuncts *at LF,* the locus of binding requirements in a minimalist theory. This assumption is doubtful, however, given that objects can license bound pronouns within adjuncts.

(38) John read every $book_i$ without reviewing it_i.

If *every book* can bind *it,* then *every book* c-commands *it* at LF. If so, it c-commands PRO as well. Regardless of whether the MDP can be made to operate in these cases, the movement approach to PRO derives the data in (36)–(37). Consider the details.

The numeration for (37) consists of the set of items {John, heard, Mary, without, entering, the, room, assorted functional categories}. We build the adjunct phrase by merging *the* with *room, entering* with *the room,* and *John* with *entering the room.* The two θ-roles of *enter* are checked by the merger of the two D/NPs. *-ing* heads its own I projection. This merges with the previously formed VP small clause. The strong feature of this I is checked by raising *John.* (Observe that this violates Procrastinate; I return to this issue after completing the proper derivation.) After the adjunct has merged with the IP, we have a structure like (39).[30]

(39) [$_{Adjunct}$ without [$_{IP}$ John [$_{I^0}$ ing [$_{VP}$ John [entering the room]]]]]

Next we build the main clause. *Mary* merges with *heard.* The internal θ-role is thereby discharged. (39) then merges with this VP, forming an adjunction structure.[31]

(40) [$_{VP/VP}$[$_{VP}$ heard Mary] [$_{Adjunct}$ without [$_{IP}$ John [$_{I^0}$ ing [$_{VP}$ John [entering the room]]]]]]

The external θ-feature of *heard* must be checked. If *John* raises, then the derivation proceeds as follows: *John* raises and discharges the external θ-role by merging with the VP of *heard.*[32] It then raises to [Spec, IP] to check its own Case, those of I, and the latter's D-features. At LF

[30] I remain agnostic on the exact status of the adjunct—that is, whether it is a complementizer or head of a PP.

[31] I represent adjunction as *VP/VP,* modifying slightly the convention in Chomsky 1995. The adjunct is adjoined at this point in the derivation to conform to the assumption that movement requires a connected tree (see Epstein 1996). It is not obvious that this assumption is required. It does not affect the discussion here in any serious way if movement between unconnected subtrees is sanctioned. See Nunes 1995, Uriagereka 1998, and footnote 32 for further discussion.

[32] Note that moving *John* from the adjunct to the higher VP is not movement to a c-commanding position. Recall that (18) sanctions such movement. This is required if the ''D-Structure'' position of the subject is not to c-command the adjunct, as appears to be empirically required. For discussion, see Hornstein 1995:chap. 8, 1998a.

Theoretically this sort of movement is important in two ways. First, it suggests that θ-marking must take place in the immediate domain of the relevant verb. In this case the external θ-marking must take place within the VP, not VP/VP. This suggests that the adjunct is adjoined, not to what eventually becomes a V′, but to the VP itself. Second, at least in cases such as these, *v* seems superfluous. If it assigned a θ-role, then we would expect the subject to c-command the

Mary raises to check accusative Case in either the outer [Spec, VP] or Agr_OP. The derivation converges with the overt structure in (41).

(41) [$_{IP}$ John [$_{I^0}$ past [$_{VP/VP}$ [$_{VP}$ John [heard Mary]] [$_{Adjunct}$ without [$_{IP}$ John [$_{I^0}$ ing [$_{VP}$ John [entering the room]]]]]]]]

This derivation requires explanation at two points. First, the movement of *John* to [Spec, ing] within the adjunct violates Procrastinate. *Mary* could have been inserted. Had it been, however, *John* could never have checked its Case features. *Mary* or a copy of *Mary* would have blocked movement out of the adjunct to a Case position in the matrix. In short, once again the MLC would prevent a convergent derivation. Thus, Procrastinate is violable at this point.

Second, *John* moves to check the external θ-feature of *heard.* Doesn't this violate the MLC? In other words, isn't *Mary* closer to this position, and shouldn't *Mary* block this movement? The MLC is not involved in this move. The reason is that *Mary* in the complement position of *heard* and *John* in [Spec, IP] of the adjunct do not c-command one another, nor does the target of movement c-command them both (see footnote 33). Thus, they are not in a ''proximity'' relation relevant for the MLC. The combination of movement from an adjunct and movement to a non-c-commanding position makes it possible for *Mary* and *John* to be equidistant from the [Spec, VP] of *heard.* Hence, moving *John* to [Spec, VP] does not violate the MLC. Furthermore, if *Mary* moves to [Spec, VP] in place of *John,* then the derivation cannot converge, for the Case features on *John* will not be checked.[33] Thus, the only convergent derivation is the one reviewed

---

adjunct. This seems to be incorrect in general, as the above references argue. A final point is that ''extension'' in the sense of Chomsky 1993 appears to ignore adjuncts. This is consistent with the position Chomsky takes in that work.

There is another, more interesting way to look at this derivation that exploits proposals made by Nunes (1995) for parasitic gap phenomena. He observes that extension can be applied to adjuncts as well if we assume that sideward movement is possible. In effect, movement is permitted out of the adjunct before the adjunct is adjoined to the VP that it modifies. Movement out of the adjunct to a θ-position inside VP ''extends'' the VP. Adjunction of the adjunct to the VP then extends the VP. In each case extension is respected. An interesting corollary is that adjuncts are not actually islands. However, extraction from an adjunct via sideward movement is *only* permitted given other natural assumptions (for details see Hornstein 1998a). If this is so, then *there* should not be able to control PRO.

(i) There arrived several men before there erupted a riot.

(ii) *There arrived several men before PRO erupting a riot.

For further discussion and details, see Nunes 1995 and Hornstein 1998a.

[33] This raises an interesting problem. What blocks *Mary* from bearing nominative Case and *John* from bearing accusative and thereby deriving (i), in which *John* checks its Case at LF?

(i) *Mary heard before John entering the room.

(ii) [$_{IP}$ Mary [$_{I^0}$ past * [$_{VP/VP}$ [$_{VP}$ Mary [heard Mary]] [$_{Adjunct}$ before [$_{IP}$ John [$_{I^0}$ ing [$_{VP}$ John [entering the room]]]]]]]

As (ii) indicates, *Mary* would have two θ-roles and *John* would check its accusative Case at LF.

This derivation is plausibly blocked by the MLC as follows. Assume that the Case-checking position for accusative is at the left edge of VP/VP, as it would be if accusative is checked in [Spec, Agr_O] or if it is checked in the outer Spec of the full VP including the adjoined clause, that is, in the outer [Spec, VP/VP] (the position of * in (ii)). This position c-commands both the adjunct and the arguments of *heard.* Consequently, this sort of movement could be blocked by the MLC if we compute distance by relativizing it to the target in cases where the potential fillers do not c-command one another. In this case distance is relevant only if the target c-commands its potential fillers. Thus, movement to a

in (39)–(41).[34] This is the desired result, for it deduces, correctly, that OC PROs inside adjuncts are necessarily controlled by subjects.[35]

6 Nonobligatory Control: The Elsewhere Case

The upshot of sections 4 and 5 is that a control module is superfluous if OC PRO is treated as a residue of movement, essentially equivalent to NP-trace. We can account for the distribution of OC PRO in terms of Case theory: it appears where Case is not checked. Typically this coincides

---

c-commanding position would trigger the MLC and related minimality restrictions, whereas sideward movement would not. (iii) would suffice for current purposes.

(iii) If a target T c-commands α, β, then α is closer to T than β is if fewer nodes intervene between T and α than between T and β.

Mary and its copies are in different domains from *John* in (ii). If the target of movement is at least as high as VP/VP, and if distance is measured as in (iii), the MLC will prevent movement out of the adjunct since *Mary* is closer to the accusative Case checker than is *John*. In other words, moving out of the adjunct is critical if the accusative Case on *heard* is to be licitly checked.

Note that something like (iii) is required if we are to make the movement analysis of PRO empirically adequate. The text relies on movement out of an adjunct. However, this is clearly not generally available. The current proposal follows Nunes (1995) in assuming that sideward movement can extract a D/NP out of an adjunct. This is how Nunes analyzes parasitic gap constructions. Such constructions also involve sideward movement to a θ-position. What needs to be blocked is extraction from an adjunct ''directly'' to a c-commanding position (e.g., [Spec, CP]). In other words, being in a c-command relation with the target appears to restrict movement in certain ways. This is what (iii) intends to capture. Another approach to these issues that gets the desired results exploits a proposal by Uriagereka (1998) regarding linearization. Uriagereka suggests that to obey the Linear Correspondence Axiom, elements that are not in a c-command relation must be linearized before they form a single subtree. Assume that this holds for adjuncts. It has the effect that adjuncts are islands, once merged with the VPs they modify. If this is correct, then LF movement *out* of the adjunct is not allowed. This derives the unacceptability of (i), as the Case on *John* remains unchecked. This is a more attractive solution than the one mooted above. However, further details require elaboration before it can be made fully operational. I leave the very complex issues involved here to further research. See Hornstein 1998a for a possible approach to these issues.

[34] This argument requires that we adopt a non-Larsonian approach to adjuncts; that is, it requires claiming that they are not like syntactic arguments. In fact, the difference between where adjuncts are merged and where complements are merged is central to explaining why the latter case (typically) manifests object control and the former, subject control.

[35] There are several issues relating to OC that I have not addressed here for reasons of space. Two are of particular interest.

The first is control in noun-complement constructions. Tim Stowell (personal communication) asks why these constructions permit control but not raising if these are essentially the same phenomenon.

(i) a. Mary's desire [PRO to win]
b. *Mary's appearance [t to win]

I address this question in Hornstein 1998a and refer the interested reader there for details.

The second, raised by the reviewer, is the asymmetry between raising and control in impersonal-SI object-preposing constructions. The wisdom concerning these constructions in Romance is reviewed in Burzio 1986:chap. 1. They are essentially the flip side of the noun-complement cases above: here raising is permitted but control is forbidden.

(ii) a. [Quei prigionieri]$_i$ risultavano [t_i essersi gia liberati t_i].
those prisoners turned out SI-to be already freed
b. *[Quei prigionieri]$_i$ vorrebbero [t_i essersi gia liberati t_i].
those prisoners would want SI-to be already freed

This contrast is addressed in Hornstein, in preparation.

with the [Spec, IP] position of nonfinite I.[36] However, OC PRO can also occur in verbs like *dress* and *shave* that have the option of not needing to check accusative Case. From the assumption that PROs are simply NP-traces—hence their inability to block *wanna* contraction and their lack of phonetic content—we can derive OC PRO's lackluster phonetic properties. In addition, we have deduced the interpretive properties of OC PROs. They require nonsplit local antecedents because they are formed by movement. They are interpreted *de se* because chains that involve PROs form compound (reflexive) predicates, that is, predicates in which an argument has two θ-roles. Finally, we have provided a theory of controllers: why object control holds in *persuade*-type verbs, why subject control is required in adjuncts headed by OC PRO. In short, reducing control to movement allows for the virtual elimination of any special control module and for the further simplification of grammatical theory along minimalist lines.

One last point. When OC is reduced to movement, the empty category in (6) (repeated here) is expected to be an NOC PRO. The reason is that in the relevant examples, movement from [Spec, IP] is prohibited.[37]

(6) a. It was believed that PRO shaving was important.
b. $John_i$ thinks that it is believed that PRO_i shaving himself is important.
c. Clinton's$_i$ campaign believes that PRO_i keeping his sex life under control is necessary for electoral success.
d. John thinks that PRO getting his resume in order is crucial and Bill does too.
e. $John_i$ told $Mary_j$ [that [[PRO_{i+j} washing themselves/each other] would be fun]].
f. The unfortunate believes that PRO getting a medal would be boring.
g. Only Churchill remembers that PRO giving the BST speech was momentous.

In the cited examples PRO is in the subject position of a subject sentence. Given that subject sentences are islands, movement from [Spec, IP] of these structures is prohibited. As a result, OC is also barred. Only NOC is permitted, as attested.

I have said very little about NOC PRO. I have silently assumed that it is identical to pro, the null pronominal found in various Romance and East Asian languages. This pro can be interpreted as a pronoun, either definite (hence similar to *he, they, she,* etc.) or indefinite (like English *one*). The latter underlies the so-called arbitrary reading. This requires assuming that pro can be licensed in English in NOC configurations. Chomsky (1993) proposes that pro is licensed in the Spec of certain Is. If pro underlies NOC readings, then *ing* and *to* can license pro, and pro can check whatever features of these Is need checking.[38] Importantly, however, the distribution of

[36] Nonfinite [+T] I, if Martin (1996) is correct. The distinction he motivates is compatible with the theory outlined here so long as one does not conclude, as he does, that Case is assigned in the Spec of such infinitival heads.

[37] Note that the landing site of movement would be a c-commanding position so that all locality restrictions should come into play given the discussion in footnote 33.

[38] This clearly needs to be stated more carefully. Small pro is not generally licensed in English. Moreover, as a reviewer correctly observes, pro does not occur freely.

(i) *I saw [pro running].

However, the ungrammaticality of (i) might well be due to a condition other than the one required to license pro in

pro is not free. OC and NOC structures do not alternate freely. Rather, NOC typically obtains when movement is prohibited—for example, from *wh*-islands, as in (42). OC and NOC are effectively in complementary distribution. This suggests that NOC is the ''elsewhere'' case.

(42) a. John told Sam how PRO to hold oneself erect at a royal ball.
b. *John told Sam PRO to hold oneself erect at a royal ball.

One way of implementing this observation would be to treat pro on a par with *do* in English. *Do*-support is regularly treated as a costly last resort operation.[39] When all other grammatical options fail to yield a convergent derivation, *do* can insert to ''support'' an otherwise deadly morphological residue. Just how to treat this ''last resort'' nature of *do*-support in the Minimalist Program is somewhat unclear.[40] Strictly speaking, *do* cannot be part of the array; if it were, sentences with and without *do* would not be comparable and so the intuition that sentences with *do* are less economical (more costly and hence to be avoided if possible) than those without it could not be theoretically grounded. Rather, *do* is a formative of the computational system of English that can be inserted in any derivation, though at a cost and, hence, only when all other relevant grammatical options have failed. A similar treatment of pro in NOC structures would yield the correct empirical results. Consider some details.

In NOC configurations, movement is impossible, as the sentences in (43) and (44) indicate.

(43) a. It is believed that Bill's/pro shaving is important.
b. *Bill's is believed that shaving is important.

(44) a. It is impossible for Bill/pro to win at roulette.
b. *Bill is impossible to win at roulette.

Consider in more detail (43a) with the pro subject in the embedded position. To license this structure, [Spec, IP] must have its features checked. There is no way to do this by moving an expression through this position to check the relevant features of I and then moving it again to check its own (e.g., Case) features, since movement from this position is prohibited. Thus, we find no OC ''PRO''s in these slots. Assume instead that pro can be inserted to meet the feature-checking requirement; in short, assume that pro needs no Case but can check the relevant features of I. Since English is not a pro-drop language, pro's features must be quite anemic. Pro then has two important properties: (a) it is able to check the requisite features of infinitival I, but (b) using pro to check such features is derivationally costly. Assumption (b) suffices to account for the fact that pro and PRO are in complementary distribution.[41] Thus, for example, (45a) cannot have

---

gerunds. Elsewhere (Hornstein 1997) I discuss the relation of local reflexives and PRO. The idea is that reflexives and PRO differ only in their Case-marking properties. If so, then (i) is blocked because (ii) is fine.

(ii) I saw myself running.

For details of this approach, see Hornstein 1997.

[39] See, for example, Chomsky 1991 and Lasnik 1995b.

[40] See Arnold 1995 for discussion.

[41] What I propose here fits English very well. Use of pro in English is clearly ''marked.'' Whether this is true in general (e.g., in pro-drop languages) is beyond the scope of this article.

an NOC interpretation, for all features can be checked without inserting pro, as the derivation in (45b) indicates.

(45) a. Mary hopes to win.
b. [$_{IP}$ Mary [$_{VP}$ Mary hopes [$_{IP}$ Mary to [$_{VP}$ Mary win]]]]

When this sort of derivation is impossible, however, pro can be inserted and the derivation saved, as in (43a). This account mimics the standard account for *do* point by point. In particular, in minimalist terms, NOC pro cannot be part of the array but is a formative used as a last option to save an otherwise doomed derivation.[42]

To sum up: OC PRO is the residue of movement and has all the characteristics of NP-trace. The only real distinction between raising and control structures is that the former involve raising a D/NP to a non-θ position whereas the latter raises expressions to θ-positions. Both raising and control chains (generally) terminate in Case positions. NOC PRO, in contrast, is simply pro, and it is licensed *at a cost* in [Spec, IP] of nonfinite CP complements.

7 Conclusion

This article has argued in favor of eliminating the control module from the grammar. In place of subtheories specially designed to account for the distribution and interpretation of PRO, the present proposal has relied on movement to account for OC and on pro to account for NOC. The price paid for thus eliminating the PRO module has been to remove the last residues of D-Structure from the grammar and to dispense with the assumption that expressions (or chains they head) are restricted to a single θ-role. Both departures from orthodoxy are required to permit movement from one θ-position to another—the minimum required if OC is to be analyzed as movement. Technically, this has also required treating θ-roles as features on predicates and θ-role assignment as a species of feature checking. It is for the reader to judge whether this price is too high for the benefits garnered. In my view, the theoretical adjustments needed to eliminate the PRO module are methodologically preferable to the theoretical stipulations they have replaced. Even if the

[42] There is some evidence that pro can function in this way in other parts of the grammar. For example, *wh*-movement is known to be able to violate weak islands if the *wh* is ''referential'' (see Aoun and Li 1993, Cinque 1991). However, there is also evidence that these instances of *wh* have not actually moved from the positions they are related to. Aoun and Li note, for example, that reconstruction effects are prohibited in such cases. Thus, whereas (i) is ambiguous, allowing a pair-list reading, (ii) is not. (ii) has only the reading on which there is at least one thing such that you wonder about whether all purchased it.

(i) What did you say that everyone bought?

(ii) ?What did you wonder whether everyone bought?

Elsewhere (Hornstein 1995) I argue that pair-list readings rely on the ability to reconstruct the moved *wh* back into its A-position—that is, to retain the A-position copy at LF rather than the one in [Spec, CP] (see Chomsky 1993). The blocking effect that *wh*-islands have on this process can be accounted for if in fact *what* does not move out of the embedded clause in (ii). That movement should be prohibited is required by the MLC (see Chomsky 1995). What occurs in (ii), then, is the binding of a null resumptive pronoun—a pro. Note that the observed degradation of (ii) can now be taken to reflect the price of using pro to save the derivation. In effect, then, pro can ''save'' a weak island violation, but at a cost.

gains were paltry, the burden of proof would be on those who favor maintaining the restrictions on θ-roles, chains, and merger that have been dispensed with.

This said, I would like to end with one set of facts that the elimination of the θ-restriction on movement has not addressed. I have said nothing about the unacceptability of simple sentences like these:

(46) a. *John saw. (meaning: John saw himself)
b. *John believes to be a fool. (meaning: John believes himself to be a fool)

The unacceptability of such sentences is the empirical basis for prohibiting movement from one θ-position to another. How can the proposed account deal with these?

I cannot provide a full answer to this question here (see Hornstein 1997). Instead, what follows is an inadequate sketch. I suggest that the sentences in (46) violate Case theory. In particular, transitive verbs like *see* and *believe* (or the *v* that is part of their structure) have an accusative Case that must be checked for the derivation to converge.[43] The movement of *John* in (46) cannot check this accusative Case and also check the nominative Case in [Spec, IP]. However, as the parenthesized paraphrases in (46) indicate, there is a grammatical alternative involving reflexives. The natural proposal is that when Case must be checked, in place of a phonetically null NP-trace what we find is a reflexive. In other words, reflexives are also the residue of movement.[44] In effect, PRO and reflexives are the same expression modulo phonetic content. Lebeaux (1985) has shown that local anaphors and OC PROs pattern almost identically with respect to the restrictions they place on their antecedents. Chomsky (1981) and Reinhart and Reuland (1993) show that locally bound reflexives have chain properties and pattern like NP-traces. A natural conclusion is that these local reflexives are just the residues of movement—how an NP-trace surfaces phonetically when there is a Case feature that needs checking. In sum, so-called locally bound reflexives are spelled-out NP-traces. This is the basic idea; the details are for another article.

References

Aoun, Joseph, Norbert Hornstein, David Lightfoot, and Amy Weinberg. 1987. Two types of locality. *Linguistic Inquiry* 18:537–577.

Aoun, Joseph, and Yen-hui Audrey Li. 1993. *The syntax of scope.* Cambridge, Mass.: MIT Press.

Arnold, Mark. 1995. Case, periphrastic *do,* and the loss of verb movement in English. Doctoral dissertation, University of Maryland, College Park.

Boškovie, Željko. 1994. D-Structure, Theta Criterion, and movement into theta positions. *Linguistic Analysis* 24:247–286.

Boškovie, Željko. 1995. Principles of economy in nonfinite complementation. Doctoral dissertation, University of Connecticut, Storrs.

Boškovie, Željko, and Daiko Takahashi. 1998. Scrambling and Last Resort. *Linguistic Inquiry* 29:347–366.

Bouchard, Denis. 1984. *On the content of empty categories.* Dordrecht: Foris.

[43] Martin (1996) assumes something similar for *believe.*

[44] See Kayne 1996 for a similar suggestion for reflexive *se* in Romance.

Burzio, Luigi. 1986. *Italian syntax.* Dordrecht: Reidel.
Chomsky, Carol. 1969. *The acquisition of syntax in children from 5 to 10*. Cambridge, Mass.: MIT Press.
Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use.* New York: Praeger.
Chomsky, Noam. 1991. Some notes on economy of derivation and representation. In *Principles and parameters in comparative grammar,* ed. Robert Freidin, 417–454. Cambridge, Mass.: MIT Press. [Also published in Noam Chomsky, *The Minimalist Program,* 129–166. Cambridge, Mass.: MIT Press (1995).]
Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger,* ed. Kenneth Hale and Samuel Jay Keyser, 1–52. Cambridge, Mass.: MIT Press. [Also published in Noam Chomsky, *The Minimalist Program,* 167–217. Cambridge, Mass.: MIT Press (1995).]
Chomsky, Noam. 1995. Categories and transformations. In *The Minimalist Program,* 219–394. Cambridge, Mass.: MIT Press.
Chomsky, Noam, and Howard Lasnik. 1993. The theory of principles and parameters. In *Syntax: An international handbook of contemporary research*, ed. Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, and Theo Vennemann, 506–569. Berlin: Walter de Gruyter. [Also published in Noam Chomsky, *The Minimalist Program,* 13–127. Cambridge, Mass.: MIT Press (1995).]
Cinque, Guglielmo. 1991. *Types of Ā-dependencies.* Cambridge, Mass.: MIT Press.
Epstein, Samuel David. 1996. Unprincipled syntax and the derivation of syntactic relations. Ms., Harvard University, Cambridge, Mass.
Fodor, Jerry A. 1975. *The language of thought.* New York: Thomas Y. Crowell.
Higginbotham, James. 1992. Reference and control. In *Control and grammar,* ed. Richard Larson, Sabine Iatridou, Utpal Lahiri, and James Higginbotham, 79–108. Dordrecht: Kluwer.
Hornstein, Norbert. 1995. *Logical Form: From GB to minimalism.* Oxford: Blackwell.
Hornstein, Norbert. 1997. Is the binding theory necessary? Ms., University of Maryland, College Park.
Hornstein, Norbert. 1998a. Adjunct control and parasitic gaps. In *University of Maryland working papers in linguistics 6,* 102–121. Department of Linguistics, University of Maryland, College Park.
Hornstein, Norbert. 1998b. Movement and chains. *Syntax* 2:99–127.
Hornstein, Norbert. In preparation. MOVE. Ms., University of Maryland, College Park.
Hornstein, Norbert, and David Lightfoot. 1987. Predication and PRO. *Language* 63:23–52.
Jaeggli, Osvaldo. 1980. Remarks on *to* contraction. *Linguistic Inquiry* 11:239–245.
Kayne, Richard. 1996. Why are reflexive clitics not canonical object clitics? Paper presented at Johns Hopkins University, Baltimore, Md., February 9.
Koster, Jan. 1984. On binding and control. *Linguistic Inquiry* 15:417–459.
Lasnik, Howard. 1995a. Last Resort and Attract F. In *Proceedings of FLSM 6,* 62–81. Indiana University Linguistics Club, Bloomington.
Lasnik, Howard. 1995b. Verbal morphology: *Syntactic Structures* meets the Minimalist Program. Ms., University of Connecticut, Storrs.
Lebeaux, David. 1985. Locality and anaphoric binding. *The Linguistic Review* 4:343–363.
Lightfoot, David W. 1976. Trace theory and twice-moved NPs. *Linguistic Inquiry* 7:559–582.
Manzini, Maria Rita. 1983. On control and control theory. *Linguistic Inquiry* 14:421–446.
Manzini, Maria Rita, and Anna Roussou. 1997. A minimalist theory of control. Ms., Università di Firenze/ University College London and University of Wales, Bangor.
Martin, Roger. 1996. A minimalist theory of PRO and control. Doctoral dissertation, University of Connecticut, Storrs.
Nunes, Jairo. 1995. The copy theory of movement and linearization of chains in the Minimalist Program. Doctoral dissertation, University of Maryland, College Park.

O'Neil, John. 1995. Out of control. In *NELS 25,* 361–371. GLSA, University of Massachusetts, Amherst.
Reinhart, Tanya, and Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24:657–720.
Rizzi, Luigi. 1986. On chain formation. In *The syntax of pronominal clitics,* ed. Hagit Borer. Syntax and Semantics 19. Orlando, Fla.: Academic Press.
Rizzi, Luigi. 1990. *Relativized Minimality.* Cambridge, Mass.: MIT Press.
Rosenbaum, Peter. 1967. *The grammar of English predicate complement constructions.* Cambridge, Mass.: MIT Press.
Rosenbaum, Peter. 1970. A principle governing deletion in English sentential complementation. In *Readings in English transformational grammar,* ed. Roderick Jacobs and Peter Rosenbaum, 20–29. Waltham, Mass.: Ginn.
Sag, Ivan. 1976. Deletion and Logical Form. Doctoral dissertation, MIT, Cambridge, Mass.
Salmon, Nathan. 1986. Reflexivity. *Notre Dame Journal of Formal Logic* 27:401–429.
Stowell, Tim. 1982. The tense of infinitives. *Linguistic Inquiry* 13:561–570.
Uriagereka, Juan. 1998. *Rhyme and reason: An introduction to minimalist syntax.* Cambridge, Mass.: MIT Press.
Williams, Edwin. 1977. Discourse and Logical Form. *Linguistic Inquiry* 8:101–139.
Williams, Edwin. 1980. Predication. *Linguistic Inquiry* 11:203–238.

Department of Linguistics
University of Maryland, College Park
Marie Mount Hall 1401
College Park, Maryland 20742

nh10@umail.umd.edu