X-bar theory, introduced in Chomsky 1970, constituted a major step toward explanatory adequacy, dramatically limiting the range of possible phrase structure rules. As with most explanatory advances, the resulting theory encounters serious descriptive problems. In the present instance, the major virtue of X-bar theory—the generalization of the phrase structure schema—immediately encountered empirical problems. For one, N and A, unlike V and P, did not seem to take NP complements:

(1)  
   a. Prove the theorem  
      b. *Proof the theorem

(2)  
   a. Resent Harry  
      b. *Resentful Harry

Since prove has a nominal and since the X-bar theory says that the complement system is general, and since prove by its semantic nature takes an NP complement, we would expect proof to take an NP as well. As is well known, this gap is completely general.

To account for this paradigmatic gap, Chomsky (1970) proposed an obligatory rule that inserts of between N or A and NP:

(3)  Of-insertion rule  

\[ \emptyset \Rightarrow \text{of} \left\{ \begin{array}{l} N \\ A \end{array} \right\} \quad \text{NP} \]

Or, in terms of the feature system of the mid-1970s:

(4)  \[ \emptyset \Rightarrow \text{of}[+N] \quad \text{NP} \]

This described the facts, but it raised two questions: Why should there be such a rule, and why should it be obligatory?

Clausal complementation created similar descriptive problems. Given underlyingly distinct finite and infinitival complements, and optional subject raising, we might have expected both (5a) and (5b) to be acceptable, but only the first of these is.
(5) a. It is likely [that Mary will solve the problem]
b. *It is likely [Mary to solve the problem]

In place of (5b), we find the raising analogue:

(6) Mary is likely to solve the problem.

Jean-Roger Vergnaud, in the personal letter to Noam Chomsky and me in early 1977 that is reprinted in chapter 1 of this book, was concerned with a class of phenomena related to this one. Vergnaud’s now very familiar basic idea was that even languages like English with very little overt case morphology pattern with richly inflected languages in providing characteristic positions in which NPs with particular cases occur. Vergnaud’s formulation, highly reminiscent of modern Case theory (particularly in its “checking” formulation), is as follows:

(7) English has three cases: Subject Case, Genitive Case, “Governed Case” (“the case of complements of verbs and of prepositions”).

(8) The restrictions on subjects of infinitivals can follow from a general filter limiting the distribution of NPs in the Governed Case.

Vergnaud offered two possible versions of the filter, both involving a structural relation discussed by Chomsky and Lasnik (1977) that is very close to “government” (which is how I will render it here):

(9) a. A structure containing an NP in the Governed Case is ungrammatical unless that NP is governed by [−N].
b. A structure containing an NP in the Governed Case is ungrammatical unless that NP is adjacent to and governed by [−N].

Chomsky (1980) built a theory of abstract case (“Case”) around Vergnaud’s insights, a fundamental step in the launching of the Government-Binding Theory. Chomsky’s initial formulation began with the following suggestion: “Suppose we think of Case as an abstract marking associated with certain constructions, a property that rarely has phonetic effects in English but must be assigned to every lexical NP” (p. 24). Thus, we have at S-structure, Case-assignment rules and a Case Filter excluding NPs lacking Case:

(10) Case assignment
a. NP is oblique when governed by P and certain marked verbs.
b. NP is objective when governed by V.
c. NP is nominative when governed by Tense.

(11) Case Filter
Lexical NPs (i.e., those with a lexical head) must have Case
*N, where N has no Case.

(12) *It was believed Mary
Compare a clausal complement, predictably acceptable, given the formulation in (11):

(13) It was believed that John won the race.

The restriction of the Case Filter to lexical (i.e., phonologically realized) NPs was, of course, crucial. One obvious indication that this is so is the acceptability of the alternative version of (12) with movement:

(14) Mary was believed to

Given trace theory, (14) does have an NP in exactly the non-Case position occupied by Mary in (12), namely, the trace of Mary. Thus, if the trace of NP-movement were subject to the filter, passive and raising structures would always be ungrammatical. An extension of this idea, developed by Chomsky (1981), is that Case is not a property of individual members of a movement A-chain but rather is a property of the entire A-chain. The Case Filter then becomes a requirement on A-chains. It is then not germane to ask whether the trace of Mary in (14) obeys the Case Filter; the chain (Mary, t) has (nominative) Case, thus satisfying the Case requirement.

However, it is still true that PRO can occur in a position where lexical NP cannot—as the subject of (certain) infinitival clauses—and this time there is no relevant chain:

(15) a. Susan tried [PRO to solve the problem]
b. *Susan tried [John to solve the problem]

We are thus still left with the lexical versus nonlexical distinction.

Chomsky (1980) noted one potential problem for this lexical versus nonlexical basis for Case theory. At S-structure, wh-NPs are typically not in any of the positions enumerated in (10), yet the resulting sentences are well-formed. Further, if the wh-NP originates in a non-Case position, the result is ill-formed. The following constitute a representative minimal pair:

(16) a. Who does it seem [__ is a nice fellow]
b. *Who does it seem [__ to be a nice fellow]
To deal with this, Chomsky resorted to a complication. Specifically, (10) is replicated internal to the rule of wh-movement:

17. Move a for wh
   a. Assign Case under (10)
   b. Addjoin a to COMP (coindexing by convention)

Now the moved wh-phrase has the Case that it received prior to moving. Given all of this, we have the following a priori plausible taxonomy:

18. NPs that obey the Case Filter
   NPs that ignore the Case Filter
   a. Lexical NPs (including Wh ones)
   b. PRO
   c. Expletives

This taxonomy makes sense since Case is treated as essentially morphophonological, and the NPs (or A-chains) that fall under its purview are morphophonologically realized—they have phonetic content.

The system seemed to work, in that it did handle contrasts like that in (16). But it was awfully cumbersome and redundant, in that all of Case assignment is replicated in the wh-movement rule. There is a rather straightforward way to eliminate the redundancy. Davis (1984) proposed that Case assignment be limited to S-structure in the general situation, but rather be an anywhere process. Whenever in the course of a derivation an NP finds itself in a Case position, Case is assigned. If that has not happened by S-structure, (11) will rule out the resulting structure, just as in Chomsky's approach. However, in the 1979 Pisa GLOW meeting presentation that became Lasnik and Freidin 1981, it was pointed out that Chomsky's (17) does not cover all of the relevant facts (and the same is true of Davis's improved version). Note first that, unsurprisingly, the paradigm in (16) is replicated with relative clauses:

19. a. I talked to the man [who [it seems ___ to be a nice fellow]]
   b. *I talked to the man [who [it seems ___ to be a nice fellow]]

There is no special difficulty here; we already know that who is subject to the Case Filter. The difficulty arises in consideration of the other forms that a relative clause can take in English. As is well known, alongside relative clauses with overt wh-operator, there are relative clauses with just the complementizer that; and those with nothing at all:

20. a. I talked to the man [who [you met ___]]
   b. I talked to the man [that [you met ___]]
   c. I talked to the man [[you met ___]]

There is thus the basis for a clear experiment. Alongside the unacceptable (19b), we should, under the Chomsky 1980 analysis, find acceptable analogues with the other two relative-clause types, those lacking relative pronouns. However, as Lasnik and Freidin (1981) pointed out, the examples are uniformly unacceptable:

21. a. *I talked to the man [who [it seems ___ to be a nice fellow]]
   b. * [that
   c. * __

The finite versions are clearly better:

22. a. I talked to the man [who [it seems ___ is a nice fellow]]
   b. [that
   c. __

Lasnik and Freidin (1981) therefore concluded that, contrary to expectation, the trace of wh-movement is subject to the Case Filter. They reformulated (11) as (23).

23. *NP, where NP is lexical or the trace of WH, and has no Case.

Chomsky (1981, 175) accepts this conclusion and gives the following formulation:

24. *[NP a] if a has no Case and a contains a phonetic matrix or is a variable.

Chomsky observes that in this respect, as in certain others, NP variables (i.e., traces of wh-movement of an NP) behave like names, for example, with respect to anaphoric possibility. As observed in Wasow 1972, and developed by Chomsky 1976, Freidin and Lasnik 1981, and Chomsky 1981, variables and names show identical effects with respect to what became Condition C of the binding theory in Chomsky 1981:

25. a. *He, said Mary likes Johni
   b. *He, said Johni likes Mary

26. a. *Who, did he, say Mary likes ti
   b. *Who, did he, say ti likes Mary

Given all of this, the PF motivation for the Case Filter is evidently gone. Chomsky proposes instead an LF motivation, attributing the basic idea to Aoun (1979). In particular, Chomsky (1981, 331 ff.) proposes reducing the Case Filter to the 8-criterion. Case renders an argument (chain) visible for 8-role assignment. A Caseless argument will violate the 8-criterion. An immediate problem arises with PRO, as in (15), which was assumed not to have Case. Thus, we wind up with the following disjunction (Chomsky’s (18) on p. 334):

27. Suppose that the position P is marked with the 8-role R and C = (a1, ..., an) is a chain. Then C is assigned R by P if and only if for some i, ai is in position P and C has Case or is headed by PRO.
Clearly the disjunction is a problem. Chomsky and Lasnik (1993, 561) eventually suggested a way of assimilating PRO to other arguments: “PRO, like other arguments, has Case, but a Case different from the familiar ones: nominative, accusative, etc. From the point of view of interpretation, we might regard PRO as a ‘minimal’ NP argument, lacking independent phonetic, referential or other properties. Accordingly, let us say that it is the sole NP that can bear null Case (though it may have other Cases as well, in nonstandard conditions that we will not review here).” Chomsky and Lasnik proposed that null Case is licensed by nonfinite Infl, but that is not quite precise. As shown by Lasnik (1993) and Martin (1996), following Stowell (1982), there is a crucial distinction between the Infl in raising and ECM constructions and that in Control configurations. It is only the latter that licenses null Case.

In addition to the conceptual simplification provided by the hypothesis that PRO is Case-marked, there is also an empirical argument presented by Chomsky and Lasnik, based on the overwhelming generalization that A-movement from a Case-checking position is barred. This prohibition has been stated in various ways over the years, including the following:

“Movement is a kind of ‘last resort.’ An NP is moved only when this is required . . . in order to escape a violation of some principle [such as] the Case filter.” (Chomsky 1986, 143)

We must “prevent a nominal phrase that has already satisfied the Case Filter from raising further to do so again in a higher position.” (Chomsky 1995b, 280)

“A visible Case feature . . . makes [a] feature bundle or constituent available for ‘A-movement.’ Once Case is checked off, no further [A]-movement is possible.” (Lasnik 1995b, 16)

“If uninterpretable features serve to implement operations, we expect that it is structural Case that enables the closest goal G to select P(G) to satisfy EPP by Merge. Thus, if structural Case has already been checked (deleted), the phrase P(G) is ‘frozen in place,’ unable to move further to satisfy EPP in a higher position. More generally, uninterpretable features render the goal active, able to implement an operation: to select a phrase for Merge (pied-piping) or to delete the probe.” (Chomsky 2000, 123)

What Chomsky and Lasnik observed is that this same prohibition seems to be in force for PRO, as illustrated by the following examples:

(28) a. *We want PRO to strike t [that the problems are insoluble]
   b. *We want PRO to seem to t [that the problems are insoluble]

Compare:

(29) a. *We want John to strike t [that the problems are insoluble]
   b. *We want John to seem to t [that the problems are insoluble]

Note that subject of the complement of want is a position where PRO is normally permitted.

(30) We want PRO to solve the problem

But it cannot arrive in that position from a Case position. This is fairly straightforward if PRO is fundamentally like John with respect to Case requirements, but mysterious otherwise.

There are still two related problems with this 8-theoretic approach to Case. First, with respect to Case, pleonastic nominal expressions seem to have exactly the same distribution as referential ones:

(31) a. *Susan tried [it to rain]
   b. *Susan tried [it to appear [that Mary is a genius]]
   c. *Susan tried [there to be an investigation]

(32) a. *It is likely [it to rain]
   b. *It is likely [it to appear [that Mary is a genius]]
   c. *It is likely [there to be an investigation]

(33) a. *My belief [it to be raining]
   b. *My belief [it to appear [that Mary is a genius]]
   c. *My belief [there to be an investigation]

In each instance, pleonastic it or there is no better in a Caseless position than a referential expression would be:

(34) a. *Susan tried [Mary to win the race]
   b. *It is likely [Mary to solve the problem]
   c. *My belief [Mary to be a genius]

Another (possibly complementary) problem is that clauses are often arguments, hence requiring 8-roles, yet they can (sometimes) appear in Caseless positions:

(35) a. My proof [that 2+2=4]
   b. It was proved [that Mary is a genius]
   c. *It was proved the theorem

With respect to pleonastics, there are two classes, according to Chomsky. First, there is weather it, which Chomsky argues is not truly a pleonastic, but is, rather, a “quasi-argument,” patterned with arguments. Chomsky’s evidence for this is that PRO, normally an argument, can be controlled by weather it, as in (36).

(36) It often rains after PRO snowing

There is further supporting evidence. The subject of try is a strongly argumental position. Yet it is reasonably acceptable, on observing a quickly darkening sky, to utter (37).

(37) It’s trying to rain.
This is in sharp contrast to the behavior of a true expletive:

(38) *It's trying to seem that Mary is a genius.

As for the apparent Case requirement of true pleonastics, Chomsky proposes that this follows from the association between these items and their "associates," forming a 0-chain. The pleonastic then requires Case, so that its clausal associate (in the instance of it) or its nominal one (in the instance of there) will be visible, under the assumption that the surface position of the associate is not a Case position. This kind of approach is often called "Case transmission." Note that this does not address (35a), with a clausal complement to a noun, and no associated expletive in a Case position. I return to this problem.

Chomsky (1981) offers a more elegant version of Case transmission (though, somewhat curiously, only for there). In particular, under the assumption that be is not a Case licensor, the proposal is that in the LF component the associate of there literally replaces it, as schematized in (39).

(39) a. There is [a man in the room] (43)
   S-structure
   b. A man is [t in the room] (44)
   LF

Now there is no need for transmission. After the movement, the associate is in its characteristic Case position—Spec of Infl. Further, the peculiar agreement displayed by existential constructions, as seen in (40), is now rationalized.\(^3\)

(40) a. There is a man in the room.
   b. There are men in the room.

The LFs show standard agreement configurations:

(41) a. A man is \(t\) in the room
   b. Men are \(t\) in the room

This has an extremely significant consequence that Chomsky did not explicitly comment on. Prior to this development, there was no observable difference between two potential theories of Case—Case assignment and Case checking. The former was presented by Chomsky (1981) (though with the acknowledgment that little was at stake). The latter is more in the spirit of Vergnaud's original proposal. But once appropriate Case configurations can be created in covert syntax, Case assignment is excluded as a possibility (at least in languages that display overt case morphology). Given the GB organization of the grammar, an LF operation cannot have PF consequences. Thus, NPs must have Case all along. The Case then must be checked—certified as appropriate—rather than assigned. Given the there-replacement analysis, checking is at the level of LF. To the extent that this analysis is successful, it provides additional reason for rejecting the morphological version of the Case Filter in favor of the Visibility approach. At this point, we have the following revised taxonomy:

(42) Items that obey the Case Filter
- Argu menstrual lexical NPs
- PRO
- WH-trace
- Argumental clauses
- (sometimes via transmission)

(43) John believes Mary to be intelligent.

In this instance, the licensing takes place in a relation not stabile in X'-terms at all, requiring the rather complicated notion "government" and additionally requiring that IP, exceptionally among maximal projections, is not a barrier to government. Assuming that raising to object position is not an available operation,\(^6\) the "exceptional" Case licensing in (43) must be reduced to a Spec-head relation. Derivation (44) provides a way to instantiate this, in terms of the Chomsky 1991 development of the Pollock 1989 split-Infl hypothesis. The movement of NP is assumed to be covert, since in PF, the ECM subject follows the verb.
Thus, the derivation is virtually identical to the *there*-replacement of Chomsky 1986. Further, once we posit raising to Spec of Agro for ECM subjects, there is no reason not to do the same for simple direct objects. Then, as suggested by Chomsky, we can reduce all structural Case licensing to one and the same configuration: Spec- (functional) head, a large step toward simplicity and symmetry in the system.

Lasnik and Saito 1991, developing observations and arguments of Postal 1974, presented evidence for this sort of raising approach. One relevant phenomenon involves the distribution of reciprocal expressions. Note that (45) is not significantly worse than (46).

(45) *The DA proved [the defendants to be guilty] during each other’s trials.

(46) *The DA accused the defendants during each other’s trials.

They both are considerably better than (47), the finite counterpart of (45).

(47) *The DA proved [that the defendants were guilty] during each other’s trials.

Given usual assumptions, the antecedent of a reciprocal must c-command the reciprocal. But an embedded subject does not c-command an adverbial in the matrix clause. This indicates that at the point in the derivation relevant to the licensing of reciprocals, or anaphors in general, the structure of (45) has changed in such a way that the position of the defendants is comparable to what it is in (46). Further, as also observed by Lasnik and Saito, the direct object in the relevant paradigms is itself possibly not high enough to c-command into an adverbial adjunct. The LF theory of Case outlined above solves both problems at once: both an object and an ECM subject raise to Spec of Agro.

Negative polarity item licensing is also known to display asymmetries characteristic of c-command determined relations. Thus, a negative subject of a simple sentence can license *any* in the object, but not vice versa:

(48) No one saw anything.

(49) *Anyone saw nothing.

Further, a negative object can, to a reasonably acceptable extent, license *any* in an adverbial:

(50) The DA accused none of the defendants during any of the trials.

Now notice that to roughly the same extent, a negative subject of an infinitival can license *any* in an adverbial attached to the higher VP.

(51) *The DA proved [none of the defendants to be guilty] during any of the trials.

This is in rather sharp contrast to a corresponding finite complement:

(52) *The DA proved [that none of the defendants were guilty] during any of the trials.

Once again, there is reason to believe that at the relevant level of representation, the subject of the infinitival complement is approximately as high in the structure as an NP complement would be, and that is higher than the base position of the NP complement.

At this point, a surprising problem arises. The model for the covert raising to Case position analysis is the Chomsky 1986 *there*-replacement. However, as pointed out by Lasnik and Saito, in ECM configurations, the associate of *there* exhibits none of the “high” behavior of argumental ECM subjects, even though under *there*-replacement, the LFs in the following pairs should be identical:

(53) a. The DA proved [two men to have been at the scene] during each other’s trials.
   b. *The DA proved [there to have been two men at the scene] during each other’s trials.

(54) a. The DA proved [no one to be at the scene] during any of the trials.
   b. *The DA proved [there to be no one at the scene] during any of the trials.
Den Dikken (1995) makes the same point for standard raising configurations, with examples like the following, involving Condition A and bound variable anaphora:

(55) a. Some applicants seem to each other to be eligible for the job.
   b. *There seem to each other to be some applicants eligible for the job.

(56) a. No applicant seems to his interviewer to be eligible for the job.
   b. *There seems to his interviewer to be no applicant eligible for the job.

There is one argument that seems to go in the other direction. Uriagereka (1988) observes that in the following example, the reciprocal can successfully take the associate as its antecedent, which is unexpected on the assumption that the position of complements is lower than that of adjuncts.

(57) There arrived two knights on each other's horses.

Uriagereka explicitly argues that LF expletive replacement, as in Chomsky 1986, is involved here. But, as shown just above, there are very strong reasons for doubting such derivations as a source of new binding possibilities. This near paradox is rather straightforwardly resolvable. Suppose that the associate of there does raise, but not all the way to the position of there. In particular, the raising to Spec of Agro already posited for transitive and simple ECM constructions could also be a possibility for existential constructions, as hinted by Lasnik and Saito 1991. This would correctly distinguish between Uriagereka's acceptable example and the unacceptable one of Lasnik and Saito and den Dikken. However, now notice, that this suggests that the Case of the associate of there is licensed independent of there itself. In fact, Belletti 1988 provides strong arguments that be (along with unaccusative verbs) is a Case licensor (licensing a Case that Belletti calls partitive). Lasnik 1992 and 1995a present additional evidence for Belletti's position. See also Boskovic 1997 and 2002.

Once we accept that there does require Case in its own right, it is reasonable to conclude that pleonastic it does as well. This then tends to implicate some version of the morphological approach to the Case Filter, as opposed to the 0-theoretic Visibility one that succeeded it. In some respects, this is a positive conclusion, since clauses are often arguments, but, as observed above, can occur in a variety of Caseless positions. One such configuration is (35b), repeated as (58).

(58) It was proved [that Mary is a genius]

Case transmission, now rejected, might have handled this. But Case transmission never had anything to say about clausal complements to nouns or adjectives, as in (59), repeated from (35a).

(59) My proof [that 2+2=4]

Visibility would incorrectly exclude (59). An idea of Stowell 1981 provides one possible way out of the problem presented by examples like (59). Stowell argued that, contrary to appearances, nouns simply do not take finite clausal complements at all. Rather, the apparent complements are actually appositives. Then, from the perspective of Visibility, it would be irrelevant for (59) that nouns are not Case licensors. However, this would not cover the full range of relevant phenomena. Stowell showed that apparent infinitival complements to nouns pass all his complement tests that the finite ones fail. From the perspective of the present discussion, then, an example like the following is still deeply problematic for Visibility:

(60) Jack's attempt [PRO to finish on time]

Clauses, whether thematic complements or not, do not need Case.

Boskovic 1995, following Kitagawa 1986, points out one interesting class of exceptions to this conclusion. In particular, the following generalization (Boskovic's (17)) obtains:

(61) Although clauses can appear in Caseless positions they need Case when they function as subjects.

Kitagawa gives the following examples (compare with That John loves Mary is surprising):

(62) a. ??I believe [that John loves Mary] to be surprising
   b. *I is likely [that John loves Mary] to be surprising
   c. *I was believed [that John loves Mary] to be surprising

To these, we could add the following contrast:

(63) a. ?My belief [that John loves Mary] is surprising]
   b. *My belief [that John loves Mary] to be surprising

Both Kitagawa and Boskovic propose that a CP can optionally be a nominal expression. One way of instantiating this is to say that the C head of CP optionally has nominal features. As Boskovic observes, this paradigm casts further doubt on a 0-theoretic approach to Case, since there is no clear thematic difference between nominal and nonnominal clauses. I will abstract away from the full details of the analysis of these facts. One possibility (definitely not the one Boskovic advocates) is that only an NP (or DP; I continue to use the label NP for ease of exposition) can satisfy the EPP, so a purely CP clause could not move to Spec of IP at all. A nominal CP would then, like other nominals, have to obey the Case Filter, yielding the result that a CP in Spec of IP will invariably require Case. A CP not in Spec of IP will "optionally" require Case, depending on whether its head is selected with nominal features or not.

The assumption of optional nominal status of CP can address another problem as well. Suppose that a Case assigner must assign its Case (i.e., a Case checker must
check its Case), a requirement sometimes referred to as the Inverse Case Filter. The problem raised by this requirement involves verbs that can freely take nominal expressions or clauses as their complement, as exemplified in (64).

(64) a. Mary believes [your story]
    b. Mary believes [that John is a genius]

If clauses don’t bear Case, (64b) would violate the Inverse Case Filter. But, as we have seen, they do not necessarily bear Case. Boskovic’s proposal is a neat way of resolving the near contradiction. At this point, though, it might be good to take a step back to examine the Inverse Case Filter. To the best of my knowledge, it was first proposed by Fukui and Speas (1986), mainly to deduce the effects of the EPP (a research goal currently still being intensely pursued). The reasoning went like this: if Spec of (finite) IP is not filled (either by movement or by insertion of an expletive), finite Inf will be unable to assign its nominative Case, under the assumption of Fukui and Speas that Inf can assign Case only to its Spec. If the Case Filter is “symmetrized,” this failure will rule out examples like (65), rendering the EPP superfluous.

(65) a. *Mary seems that Mary is a genius
    b. *Mary has been a man in the room

EPP in a nonfinite clause does not completely fall under such an analysis. For ECM contexts, Fukui and Speas propose that filling the Spec of IP is necessary so that the matrix verb can assign its accusative Case (under government; if the embedded subject, or passivized object, were VP internal, VP would act as a barrier to government). This is already not without difficulties, as discussed above. Successive cyclic A-movement, as in (66), is still more difficult.

(66) Mary is believed [to be likely [t’ to be chosen]]

The movement to matrix subject position can be assumed to be Case-driven, but not the movements to t’ or t". If we assume that the sole driving force for A-movement is Case, we are led to a global conception of economy of derivation, as discussed by Chomsky (1995a). In particular, we would be led to the conclusion that a specific instance of movement need not immediately result in the checking of a feature. Rather, it need only be a necessary step in a sequence of movements that ultimately result in the required checking configuration. For Chomsky (1995a) the intermediate steps in a successive cyclic A-movement derivation are licensed in this way. A more local, Markovian, conception of economy implicates the EPP. If each Inf, finite or not, has an EPP requirement, then (assuming a bottom-up structure-building mechanism) as soon as Inf is introduced into the structure, that EPP requirement will have to be satisfied. No “look-ahead” will be needed. And now notice that if the EPP is required completely independent of the Inverse Case Filter, it is the latter, rather than the former, that might be rendered superfluous. This is a significant enough issue that it is worth examining whether A-movement is, in fact, successive cyclic. Chomsky (1981, 44–45) presents an empirical argument for intermediate A-traces, hence, for successive cyclic A-movement. He observes that (67) is acceptable, indicating that Condition A is satisfied.

(67) They are likely [t’ to appear to each other [t to be happy]]

Based on this, he argues for intermediate traces (hence, for successive cyclic movement). His line of reasoning is as follows: “The GF-Ø filled by medial traces such as t’…may…be relevant to LF; for example in the sentence [(67)],…where the medial trace serves as the antecedent of each other, which requires an antecedent in the same clause in such sentences in accordance with binding theory.” Interestingly, though, it does not seem that the antecedent must be in the same clause, given the binding theory in Chomsky 1981 (or those in Chomsky 1973 and 1986 for that matter). Consider the Chomsky 1981 formulation:

(68) β is a governing category for α if and only if β is the minimal category containing α, a governor of α, and a SUBJECT accessible to α.

(69) SUBJUNCT = AGR in a finite clause; NP of S in an infinitival;…

(70) γ is accessible to α if α is in the c-command domain of γ and…

If there is a trace in the intermediate clause then that clause is the governing category (GC) of each other. But that is no argument that there is such a trace. Suppose there were none. Then the intermediate clause, lacking a SUBJECT, would not be the GC, and Condition A would, correctly, be satisfied with they as the binder of each other. Thus, there is no argument here for successive cyclicity of A-movement.8

I have just shown that Chomsky’s argument for successive cyclicity does not go through on standard formulations of the binding domain for anaphors. Chomsky’s argument relied on a clause-mate condition, but those formulations do not require that an antecedent of an anaphor be a clause-mate in the crucial cases. However, in Lasnik 2002 and 2003, I argued that clause-mate is, in fact, the correct characterization, as argued much earlier by Postal 1966 and 1974, and generally assumed until Chomsky 1973. My argument was initially based on a very interesting verb-particle construction first discussed by Kayne 1985 and later analyzed by Johnson 1991 in terms relevant to the present discussion. Johnson provided an insightful account of examples like (71) involving overt raising of the ECM subject John.

(71) Mary made John out to be a fool.

Both Kayne and Johnson convincingly treat (71) as an infinitival counterpart of (72).
(72) Mary made out that John is a fool.
Very plausibly, John in (71) has overtly raised into the matrix clause. Now consider that for many speakers, the raising seen in (71) is optional. For those speakers, (73) is an acceptable alternative to (71).
(73) Mary made out John to be a fool.
This time, plausibly John has remained in the embedded clause, and in Lasnik 2002 and 2003, I provided evidence that this is so. Now note that an anaphor on the upstream side of out is much better than one on the downstream side, with antecedent in the matrix clause:
(74) a. Jack made himself out to be honest.
b. *Jack made out himself to be honest.
(75) a. They made each other out to be dishonest.
b. *They made out each other to be dishonest.
None of the standard formulations of the notion governing category predict the disparity between the (a) and (b) versions of these examples, but a clause-mate formulation would. In each of the (b) examples, the anaphor is separated from its antecedent by the boundary of the infinitival clause, while in the (a) examples, no clause boundary of any sort intervenes.
None of the standard formulations of the notion governing category predict the disparity between the (a) and (b) versions of these examples, but a clause-mate formulation would. In each of the (b) examples, the anaphor is separated from its antecedent by the boundary of the infinitival clause, while in the (a) examples, no clause boundary of any sort intervenes. It must be acknowledged that there is an interfering factor: Verb-particle-NP order with reflexives and reciprocals is somewhat degraded even with only one clause, as in (76)–(77).
(76) ?Jack called up himself.
(77) ?They called up each other.
However, the deviance is less severe than that seen in ECM instances like (74b)–(75b). Thus, the argument for the clause-mate condition stands.
Given the clause-mate requirement on anaphors, examples like the following, attributed to Danny Fox, via David Pesetsky, in Castillo, Drury, and Grohmann 1999, argue for successive cyclic A-movement:
(78) John appears to Mary [to seem to himself/*herself [to be the best candidate]]
In the absence of successive cyclic movement, himself would incorrectly be predicted to violate Condition A, and, on fairly standard assumptions, herself to be in conformity.
A similar, though somewhat less direct, argument can be constructed based on Condition B. Examples (79)–(80) display familiar Condition B effects.
(79) *JohnÂ¿ injured him;
(80) *JohnÂ¿ believes him, to be a genius
As would be expected, (79) remains unacceptable under VP ellipsis:
(81) *Mary injured him, and John, did too
What is not expected is that (80) is substantially improved under VP ellipsis:
(82) ?Mary believes him, to be a genius and John, does too
How is it that deletion, which I take to be a PF process following Ross 1969, Chomsky and Lasnik 1993, and Lasnik 1999b, among others, remediates a presumably semantic violation? Suppose that in addition to their properties with respect to anaphora, the pronouns under consideration also have a morphosyntactic requirement, in particular that as weak pronouns they must cleftize onto the verb, as suggested by Oehrle 1976, based on data like that presented by Chomsky 1955:
(83) The detective brought John in.
(84) The detective brought in John.
(85) The detective brought him in.
(86) *The detective brought in him.
If the relevant structural configuration for Condition B is based on the notion clause-mate, an account of the ellipsis paradigm presents itself. In particular, it is not actually the Condition B violation (presumably an LF effect) in (80) that is repaired by ellipsis in (82). Rather, it is failure to cleftize (a PF violation) that is repaired (by PF deletion). And without cleftization, the pronoun can remain in the lower clause (if, as argued in Lasnik 1999a and 2001, “subject raising” in these constructions is generally optional). On the other hand, in (81), the pronoun and its antecedent are clause-mates independent of cleftization, so there is no possibility of “repair” of the Condition B violation by ellipsis. Given the clause-mate character of Condition B, (87) constitutes evidence for successive cyclic movement, because John must have moved through the intermediate Spec of IP to yield the observed obviation. John in its surface position is not a clause-mate of him.
(87) *JohnÂ¿ is believed [to seem to him, to be a genius]
Given this motivation for the EPP, I will remain agnostic about the Inverse Case Filter. It is worth observing that under the Agree theory of Chomsky (2001), neither the Inverse Case Filter nor even the classic Case Filter ever drives movement. Whatever checking by a head needs to take place can, in fact must, take place prior to movement of an XP to the specifier position of that head, since Agree demands that the Case licensing head (the Probe) c-command the Case licensee (the Goal). Phrasal movement is, then, invariably driven by the EPP. I do continue to assume some version of the classic Case Filter, since its effects are not entirely deducible
from the EPP, nor from anything else, as far as I can tell. Thus, the original question is still germane: Exactly what are the items that must obey the filter? The discussion here has led to the following empirically motivated taxonomy, admittedly problematic in that it does not make much sense, though very simple: NP A-chains obey the Case Filter.

(88) Items that obey the Case Filter
- Argumental lexical NPs
- PRO
- WH-traces
- Expletives
- Clauses (nominal)

(89) Items that ignore the Case Filter
- Clauses (nonnominal)
- (NP traces)
- Everything else

The second fundamental question about the Case Filter also still persists: Exactly where does it have to be satisfied? Is it an LF requirement or a PF one? Conceptually, either can be motivated. But, as seen above, in both instances, the motivations are not without difficulties. I have no definitive way to resolve this issue at the moment. I will, however, end with a new argument suggestive of a PF approach. Postal (1974) discovered a class of verbs that take infinitival complements, but where the subject must (to put it in more modern terminology) be an A'-trace. *Allege* is one of about two dozen that he lists. With these verbs, we find the unacceptable (89) rescued by such operations as topicalization (90).

(90) *I alleged John to be a fool.
I alleged to be a fool.

Kayne (1984), followed by Boskovic (1997), takes (89) to be a violation of the Case Filter. For Kayne and for Boskovic, A'-movement provides a way of satisfying the Case Filter violation. Unlike standard ECM verbs, *allege*-type verbs are somehow not powerful enough in their Case-checking ability to reach all the way down into the spec of the IP in their complement; the first step of A'-movement brings the relevant DP into close enough proximity for its Case to be checked. The accounts of Kayne and Boskovic differ in many details, but here I will abstract away from the details entirely, and just assume that they are correct that the Case Filter is implicated and that A'-bar movement somehow provides a way that it can be satisfied. Consider now the sequence of sentences in (91).

(91) a. John, I alleged to be a fool.
b. Mary did [allege John-to be a fool].

In (91a), we have the aforementioned satisfaction of the Case Filter via A'-movement. (91b) is more interesting. The source must have been as indicated, since it is obvious that topicalization did not take place. (If it had, *John* would be far out-of-the-deleted VP and would be pronounced.) The question is why *John* in (91b), different from its nonelliptical source, is not ruled out by the Case Filter. Along the lines of Merchant 2001, Lasnik 1995c, and Lasnik 1999b, among others, I suggest that we have here an instance of repair by ellipsis. In particular, a PF process (deletion) is repairing a violation, indicating that the violation is (or would have been) a PF one. Thus, we have some rather surprising evidence that the Case Filter is, in fact, a PF requirement. What it means for items with no phonetic content (PRO, WH-trace) to have to obey a PF requirement is a question I will have to leave for future consideration.

Notes

I am very pleased to be able to contribute a discussion of Case theory to a book celebrating Jean-Roger Vergnaud’s many contributions to linguistic theory. The importance of Case for syntactic analysis was, of course, entirely Jean-Roger’s discovery. And it is hard to imagine the principles and parameters framework ever being developed without this crucial insight. I would also like to take this opportunity to thank Zeljko Boskovic, Cedric Boeckx, Bob Freidin, and Noam Chomsky for hundreds of hours of discussion with me over the years (or decades) of Case theory and many related matters. Without those discussions, my research would have come to a dead end on many occasions. Finally, I would like to thank Tomohiro Fujii, Bob Freidin, and two reviewers for many penetrating suggestions for improvement of an earlier draft.

1. An alternative conception, more like Vergnaud’s, assumes that all NPs are “born” with Case and the Case Filter excludes mismatch between the Case an NP already has and the position in which the NP finds itself. I will return to potential ways of teasing these two conceptions apart. In passing, I note that Chomsky’s taxonomy of Case types is a bit of a departure from the classical terminology, under which all cases other than nominative are classified as oblique. I am grateful to Scott Olson for reminding me of this.

2. As already demonstrated by Jespersen (1927), the *that* introducing a relative clause is, indeed, a complementizer and not a “relative pronoun.”

3. The agreement patterns are not always this simple. See Sobin 1997 for an important discussion of agreement when the associate is a coordinate structure.

4. At this point, I intend the term licensing as neutral between a mechanism of Case assignment to a previously Caseless NP (as in Chomsky 1980, 1981) and one of “checking,” under which an NP is initially inserted into a structure with a Case feature whose appropriateness to its position is later certified. As discussed just above, under some circumstances, an NP is not in a Case position until the LF component, arguing for a checking approach.


8. Bob Freidin points out that I am here assuming that the completely empty subject of the intermediate clause would not count as any kind of SUBJECT. Note that in later developments of the theory, there would be literally no subject at all unless something actually moved into that position.

9. Postal (1974) provides independent evidence that pronominal ECM subjects show the same cliticization behavior as pronominal objects.

10. Boskovic (2002) provides a new argument for the Inverse Case Filter, based on (i) and (ii):

(i) *Mary loves here/there.

(ii) a. Mary loves it here/there.
   b. Mary loves this/that place.

Boskovic reasons that the examples in (i) are perfectly coherent (as demonstrated by (ii)), and are bad just because here and there cannot bear Case. However, this phenomenon is much more limited than would be expected. The following examples contrast minimally with Boskovic's:

(iii) a. Mary found/discussed this place.
   b. *Mary found/discussed here.
   c. *(*)Mary found/discussed it here.

[Good only on the irrelevant reading where it denotes some object.]

(iv) a. I talked about this place.
   b. *(*)I talked about here.
   c. *(*)I talked about it here.

[Good only on the irrelevant reading where it denotes some object.]

I leave for future research further investigation of (i) and (ii), but note in passing that Lydia Grebenyova plausibly suggests relating them to

(v) a. I love it when you sing.
   b. I love when you sing.

11. Bob Freidin informs me that for some speakers, allege behaves just like consider, licensing ECM even when the subject of the infinitival clause has not undergone A'-movement. Needless to say, for such speakers the following argument does not go through. I suspect, though, that for all speakers, at least some verbs on Postal's list pattern as he reports.

12. Two reviewers each suggest alternative sources, which I will briefly consider here. The first alternative relies on the proposal of Merchant 2001 that there is no formal identity requirement for ellipsis, just a purely semantic one. Under this approach to ellipsis, the elided material in (91) could be allege that John was a fool, rather than the indicated infinitival. This would then raise no issues of Case licensing. However, there is at least some reason to believe that formal identity is at least to some degree relevant in licensing ellipsis. One such reason is provided by Merchant himself. Active-passive pairs typically do not alternate:

(i) *Someone shot Ben, but I don’t know by whom.

In the absence of any formal identity condition, it is not immediately clear why ellipsis is not possible here. Merchant proposes that the subject of the active transitive induces relevant entailments that the by-phrase does not. This might turn out to be the right direction, but as it stands, it is just a promissory note. There are two other residues of formal identity that could be mentioned here. One is the fact that for many speakers, sloppy identity is disfavored if there is a mismatch of agreement features:

(ii) ??Mary washed her car and John did [wash-his-car] too

The second is a restriction on VP ellipsis with forms of be discussed in Warner 1986 and analyzed in terms relevant to the present discussion in Lasnik 1995d. As illustrated in (iii), finite forms of be cannot anteced the infinitive.

(iii) *Mary is a doctor and John will [be-a-deezer] too

Here again it is hard to see how any semantic identity could be at issue.

The second alternative is based on the observation that while the allege class of verbs do not license Case on full DPs, they do, for many speakers, on weak pronouns (perhaps via incorporation):

(iv) I alleged *John?him to be a fool.

The elided material in (91) could then be allege him to be a fool, once again obviating any Case difficulty even without ellipsis. And even accounts of ellipsis demanding formal identity necessarily allow this kind of "vehicle change" in the sense of Fiengo and May 1994. However, Tomo Fujii shows that this kind of account cannot cover all of the relevant data. Consider (v):

(v) His, mother, John, alleged to be beautiful. Bill did too:

Fujii observes that sloppy identity is possible here, unexpected if the elided material were simply allege her to be beautiful.

Finally, I note that one of the A'-ECM constructions discussed by Kayne is not even potentially amenable to a vehicle change to pronoun analysis. In (vi), a pronoun is no better than a full DP.

(vi) *I assure you John/him to be the best candidate.

Compare (vii).

(vii) I assure you that John is the best candidate.

Here, as with Postal's examples, A'-movement rescues the violation:

(viii) *John, I assure you to be the best candidate.

Significantly, the VP ellipsis pattern with allege can also be replicated:

(ix) ?John, I assure you to be the best candidate. Mary will too.

References


