The articles in this book are concerned with the treatment of various anaphoric phenomena. The framework is that first introduced by Chomsky (1973) and successively revised and refined throughout the '70's and '80's. In this chapter, I will present a highly selective history of one line of investigation during that period in what has come to be known as Binding Theory (BT). I will be concerned with how Binding Theory as explicated by Chomsky has changed, but also with how it has remained the same (sometimes despite appearances to the contrary); with how changes in other modules of the theory have precipitated changes in BT; and with general theoretical concerns, for example, the rule of the desire to eliminate redundancy in the theory. Where appropriate, I will indicate how the articles comprising this book fit into the developments discussed. All references to Chapters will be to those in this book. In the interests of clarity, these articles will be referred to by their chapter numbers as well as by their original dates of publication, thus Chapter 4 [Lasnik (1976)]. It should be kept in mind that publication dates can be misleading as to history, since the large majority of the books and articles referred to in this chapter were in fairly general circulation in manuscript form long before they were actually published. It should also be kept in mind that because of the limited scope of the discussion, a number of important approaches to anaphora are ignored, including the influential Linking theory of Higginbotham (1983, 1985). Even within the approach discussed, questions of 'reconstruction' and of levels of representation relevant to BT will not be addressed. See Barss (1986) for discussion.

A number of new familiar themes were first introduced in Chomsky (1973). Perhaps most important is that there is one domain relevant to both disjoint reference effects and (indirectly) anaphoric dependence. The domain is characterized in terms of two constraints, the Tensed S Condition (TSC) and the Specified Subject Condition (SSC). These are presented, to a first approximation, in (1b) and (1a) respectively:

\begin{enumerate}
\item No rule can involve X, Y in the structure
\[ \ldots X \ldots \{ \ldots Z \ldots WYV \ldots \} \ldots \]
where (a) Z is the specified subject to WYV
or (b) α is a tensed S
\end{enumerate}

(1b) straightforwardly excludes a passive such as (2a), while allowing (2b):

\begin{enumerate}
\item a. *John is believed [is happy]
\item b. John is believed [to be happy]
\end{enumerate}
The embedded finite clause counts as \( \alpha \) for (1b). Note in passing that it is the movement itself that is prevented by \( \alpha \). While traces were introduced in Chomsky (1973), their treatment as anaphors was a later development, which I will discuss below. In fact, Chomsky (1973) had no 'interpretive' processes establishing relations between anaphors and antecedents. Reflexives were not discussed at all. Further, as just noted, the trace of NP movement was not treated as an anaphor. Reciprocal 'each other' was also not treated in such terms. Here, too, the analysis was in terms of a syntactic transformation whose operation was constrained by (1). Following Dougherty (1970), Chomsky assumed a rule of *each*-Movement ultimately relating a more abstract structure like (3) to a more superficial one like (4).

(3) The candidates each hated the others

(4) The candidates hated each other

As in the case of passive, it is the movement that is constrained. In a footnote, Chomsky mentions the possibility, suggested by Jackendoff (1969), that the relevant process is a 'rule of interpretation' rather than one of movement. In Chapter 2 [Fiengo and Lasnik (1973)], that position is defended and discussed in detail.

While Chomsky (1973) had no direct coreference device, he did have an explicit non-coreference device, one whose effects are very similar to the ultimate effects of Condition B of Chomsky (1981). This was his RI (rule of interpretation) which

... applying to the structure NP-V-NP (among others) seeks to interpret the two NPs as nonintersecting in reference, and where this is impossible (as in the case of first and second person pronouns ...), it assigns "strangeness," marking the sentence with *.

Among the examples Chomsky presented are the following:

(5) *I saw me

(6) *We watched me leaving (in the mirror)

(7) He saw him

As noted, (5) and (6) are excluded by RI. (7), on the other hand, is allowed, but its interpretation is restricted in a particular way. Chomsky argued that RI, like passive and *each*-Movement, is constrained by (1) above. Thus, as long as the two NPs are separated by \( \alpha \), the rule seeking to interpret them as nonintersecting will be blocked. (8) does not display the "strangeness" of (5), nor does (9) display the strangeness of (6):

(8) I think I will win

(9) We think I will win

Note that the effect of (1) on RI is to allow additional sentences, or additional readings of sentences. This is the opposite of the effect of (1) on passive or *each*-Movement. This follows directly from the filter-like property of the former vis-à-vis the latter.

Thus far, the examples of the operation of (1) have all involved (1b), the Tensed S Condition. (1a), the Specified Subject Condition, is illustrated below. (10) shows how (1a) blocks *each*-Movement. In (10a), \( \alpha \) is the complement infinitive clause, and in (10b), it is the object NP. In both examples, Mary is Z, the specified subject.

(10) a. *The men believe [Mary to like each other]
   b. *The men like [Mary's pictures of each other]

(11) displays the same paradigm, but with RI the relevant rule. The men and *them* may be coreferential.

(11) a. The men believe [Mary to like *them*]
   b. The men like [Mary's pictures of *them*]

I turn now to the notion 'specified subject' in (1a). In particular, under what circumstances might a subject not count as specified? Chomsky's answer is that the notion is determined relative to choice of X: Y is a specified subject with respect to X if it is not 'controlled' by X or a category containing X. Control here is both the relation between what is now called PRO and its antecedent (which is currently still called Control), and also the relation between a trace and the moved item that it is the trace of. A lexical Y is not controlled at all in this sense, hence is not controlled by X. Such a Y is thus always a specified subject, and this is taken as the core case. A non-lexical Y might or might not count as a specified subject, depending on what its controller is. In (12), PRO counts as a specified subject with respect to the matrix subject, but not the matrix object. This is because PRO is controlled by the latter and not by the former.

(12) a. *We persuaded Bill [PRO to kill each other]
   b. Bill persuaded us [PRO to kill each other]

Thus, SSC blocks *each*-Movement from the first person plural pronoun only in (12a). The pattern in (13) is precisely the reverse, in accord with the fact that *promise* takes subject control, rather than object control.

(13) a. We promised Bill [PRO to kill each other]
   b. *Bill promised us [PRO to kill each other]

The treatment of Wh-Movement, and of the interactions between this process and others, played a significant role in the theory. First, the contrast between (14a) and (14b) was treated in terms of (1), with (14b) excluded by the SSC.
overlooked at the time, and for some time subsequently. Thus, in Chomsky (1976), 'specified' is still assumed to play a crucial role in the statement of the conditions.

Beginning with the framework of Chomsky (1973), we have now seen evidence for two modifications. In particular, it has been suggested that Wh-movement does not fall under the constraints in (1). In addition, it has been argued that all subjects are specified subjects. Both of these modifications lead to simplifications in (1), and have empirical advantages as well, as seen above and in Chapter 5 [Freidin and Lasnik (1981)]. The first modification raises the question of whether it is plausible that certain rules should be exempt from the conditions. Chomsky himself claims, in passing, that there is a rule that does not fall under (1).

Notice that one rule that obviously does not satisfy the condition is Coreference Assignment (however it is formulated). Thus the pronoun can be anaphoric in John said that he would leave, for example. The same rule also applies within coordinate structures (for example, John said that he and Bill would leave) and others that block various types of rules.

On the other hand, in Chapter 4 [Lasnik (1976)], it is argued at some length that there is, in fact, no rule of coreference assignment responsible for the coreferential interpretation of pronouns in examples of the sort mentioned by Chomsky in the passage quoted just above. Based in part on the conclusion of this argument, Chomsky (1976) rejects his earlier claim. He seems to concede that if there were a rule associating he with John in the examples mentioned, this might undermine the approach developed in Chomsky (1973) and elaborated in Chomsky (1976). However, referring to Lasnik (1976), he indicates that

The rule of anaphora involved in the (normal but not obligatory) interpretation of [John thought that he would win] should in principle be exempt from the conditions on sentence-grammar, since it is no rule of sentence grammar at all.

This conclusion is not entirely unproblematic, however. While no sentence-grammar rule of coreference will be involved in the interpretation of pronouns, given the approach of Chapter 4 [Lasnik (1976)], there is, crucially, a sentence-grammar rule of non-coreference (or, more generally, disjoint reference) involved in the analysis presented. Consider (32):

\[
(32) \quad \text{He thought that John would win}
\]

The relevant property here is strongly reminiscent of that seen in such an RI example as (33):

\[
(33) \quad \text{John likes him}
\]

In particular, in both (32) and (33), we have two NPs that are assigned disjoint reference. In Chomsky (1973), as discussed earlier, the relevant rule for (33) was RI. In Chomsky (1976), essentially the same rule persists, but with a new name, 'Disjoint Reference (DR). This rule "assigns disjoint reference to a pair (NP, pronoun) ..." Crucially, of course, Chomsky's DR is constrained by TSC and SSC. Yet the rule implicated in (32), whose semantic effects are apparently identical to those of DR, clearly does not conform to the two constraints, as is immediately evident in (32) and (34):

\[
(34) \quad \text{He believes Mary to like John}
\]

In (32), the relevant interpretive rule, assigning disjoint reference, operates in apparent violation of TSC; and in (34), it operates in apparent violation of SSC. Thus, the problem noted by Chomsky (1973) still remains; only the rule has changed. In fact, this problem remained until the more radical reformulation of BT in Chomsky (1980), which will be discussed below.

In Chomsky (1976), we do find one rather significant departure from Chomsky (1973), namely, the conjecture that TSC and SSC are solely constraints on anaphoric relations of a particular type, rather than on the operation of transformations. Reciprocals are now treated in terms of a reciprocal rule which "assigns an appropriate sense to sentences of the form NP ... each other", as in Chapter 2 [Fiengo and Lasnik (1973)]. Further, it is proposed that the relation between an NP that has undergone NP-Movement and the position from which it moved is one of bound anaphora, as proposed by Fiengo (1974). As such, it would naturally fall under the same conditions as reciprocal interpretation. Chomsky thus tentatively concludes that

... we can regard SSC, in such cases, as a condition on surface structures applying quite generally to anaphora (hence to the NP-trace relation), rather than a condition on transformations.

Curiously, though, an equivocation remains concerning Wh-movement. It is suggested that this rule "seems superficially to violate ... SSC, and [TSC]." However, since the rule is (successive) cyclic and "movement is permitted from complementizer position in a tensed-S", there is no actual violation. As seen above, technically, the conditions can be made consistent with the properties of Wh-movement in this way. However, complications are required, with no obvious concomitant benefit. Further, if TSC and SSC are to be regarded strictly as constraints on anaphoric relations, as suggested above, there is no a priori reason to even expect them to be relevant to Wh-movement. This is so since, as Chomsky argues, the trace of Wh-movement does not have the properties of an anaphor, but rather, those of a name. Strong crossover is handled in just these terms, essentially following Wasow (1972). Thus, Chomsky compares (35a—c) with (36a—c) respectively.
(35) a. Who t said Mary kissed him
b. Who did he say Mary kissed t
c. Who t said he kissed Mary

(36) a. John said Mary kissed him
b. He said Mary kissed John
c. John said he kissed Mary

Strong crossover, the fact that he in (35b) cannot be understood as a variable bound by Who, thus reduces to just the disjoint reference phenomenon evidenced in (36b), but only if the trace of Wh-movement is treated as a name. However, as will be discussed shortly, this line of reasoning did not entirely settle the issue.

Weak crossover, as in (37), for example, was also investigated in Chomsky (1976).

(37) a. *Who_1 does his_1 mother love t_1
   b. *His_1 mother loves everyone_1

The core of Chomsky’s analysis of (37a) is a constraint on the relationship between a pronoun and a variable antecedent: A variable cannot be the antecedent of a pronoun to its left. This analysis extends to (37b) on the assumption that the constraint is applicable to Logical Form, where the surface structure position of everyone will be occupied by a variable bound by everyone. A general, though brief, account of the circumstances under which pronouns can function as bound variables is presented in Chapter 4 [Lasnik (1976)]. There the structural requirement relevant to weak crossover, assumed to be precedence and command, is argued to be involved in the phenomenon of ‘slippery identity’ as well, indicating that the latter also involves variable binding. See Reinhart (1983) for a detailed discussion of this correlation.

Chomsky (1980) presents the first major revision of modern binding theory. It is here that the notions ‘bound’ and ‘free’ are introduced, with their still standard definitions, as in (38a, b):

(38) a. An anaphor α is bound in β if there is a category c-commanding it and coindexed with it in β.
b. Otherwise, α is free in β.

The SSC and TSC (actually called P(repositional) I(slund) C(ondition) since Chomsky (1976)), are restated initially as (39):

(39) If α is an anaphor in the domain of the tense or the subject of β, β minimal, then α cannot be free in β, β = NP or S’.

(39) is, of course, quite similar to Condition A of Chomsky (1981).

Chomsky calls (39) the “Opacity Condition”, suggesting that Tense and Subject are “operators” making certain domains opaque to binding from without. He indicates that the Opacity Condition differs in important respects from earlier formulations. First, what was merely a conjecture previously is now fully implemented: here reciprocals, reflexives, and NP trace are all treated as anaphors; (39) does not directly constrain the operation of any transformation. Conceptually, this seems a desirable simplification, though, as discussed in some detail in Chapter 7 [Lasnik (1985)], rather surprisingly there are certain constraints on NP movement that cannot obviously be handled by conditions on the distribution of anaphors.

Second, the notion ‘specified’ is finally eliminated from the SSC, along the lines suggested above. As noted, this has both conceptual and empirical benefits. In particular, ‘empty categories’ are now full-fledged syntactic entities, serving as anaphors, as specified subjects, and, finally, as antecedents.

The basic Opacity Condition in (39) is modified in certain respects in Chomsky (1980). The first modification is motivated by a certain ‘redundancy’ present in (39), and, in fact, present in the earlier versions of the conditions as well. Consider (40):

(40) They told me [what I gave each other]

The reciprocal each other must find an antecedent; further, the antecedent must be plural, hence, not I in (40). (39) correctly prevents They from being the antecedent. However, as Chomsky notes, (39) prevents this in two different ways. each other is free in the domain of tense in the lower clause (the TSC/PIC); it is also free in the domain of the lower subject I (the SSC). More generally, PIC excludes free anaphors from both subject position and non-subject position of a finite clause, while SSC excludes free anaphors from non-subject position of both finite and non-finite clauses (and of NPs, as well). The conditions overlap with respect to non-subject of a finite clause. Chomsky proposes eliminating this redundancy by restricting the PIC to the subject of a finite clause. This is accomplished by limiting (39) to SSC effects, and adding an additional condition involving subjects of finite clauses. The revision of (39) is presented in (41), and the new condition is shown in (42):

(41) If α is in the domain of the subject of β, β minimal, then α cannot be free in β.

(42) A nominative anaphor cannot be free in S'.

Chomsky calls (42) the NOMINATIVE ISLAND CONDITION (NIC). Strictly speaking, this modification does not quite eliminate all redundancy
between PIC and SSC, for there is still a small residue of redundancy between NIC and Opacity. In (43), the nominative anaphor each other is free in S', and is also free in the domain of a subject, Mary:

(43) *The men think that Mary said [that each other would win]

But certainly the central instances of redundancy between the two conditions are now removed.

In addition to the advantage of eliminating a redundancy (or most of one, as it turns out), Chomsky argues that the reformulation in (41)–(42) is of empirical benefit also. Examples such as (44) constituted problems for TSC/PIC:

(44) They expected that pictures of each other (each other's pictures) would be on sale

Compare (45):

(45) *They expected that each other would be there

(44), unlike (45), is grammatical, but the PIC would incorrectly prohibit They from serving as the antecedent of each other in the former. In earlier work, Chomsky had suggested that PIC should be constrained by Subjacency. Since each other is not subjacent to They in (44), PIC would not take effect. But now there is a simpler account available: each other is nominative in (45) but not in (44). Hence, only (45) is constrained by NIC (and, of course, neither example falls under SSC). And under this simpler account, Subjacency can be construed as strictly a property of movement rules.

Note that this discussion relies on something of an equivocation. Thus far, RI effects have not been considered; the disjoint reference in, say, (46) is not yet incorporated into the Chomsky (1980) framework.

(46) a. *I/We like me  
   b. They like them

This omission is of potential significance in the treatment of (44), since not just an anaphor, but also a bound pronoun, is possible in the position under investigation. Consider (47):

(47) They expected that pictures of them (their pictures) would be on sale

The pronoun in the 'picture NP' in these examples is free to corefer with the matrix subject. Thus, it will ultimately be crucial to keep RI from applying in certain contexts where bound anaphors are possible. Below, this point will be considered further, when the treatment of disjoint reference in Chomsky (1980) is presented. At present, it should be kept in mind that essentially all versions of Binding Theory from Chomsky (1973) through Chomsky (1981) are designed to capture the complementarity of anaphors and bound pronouns. This is a virtue of these versions of the theory, since such complementarity overwhelmingly obtains. But it is a defect as well, since, as just seen, the complementarity is not total.

There are two other points that should be noted, these concerning special properties of traces. First, since NP trace is an anaphor, it must conform to NIC and Opacity. And since it is created by movement, it must conform to Subjacency. Thus, Chomsky observes that (48) is not ruled out by either NIC or Opacity, since the trace is in a position permitting an anaphor, as seen in (44) above.

(48) *The men were expected that pictures of t were on sale

But (48) (or its derivation) does violate Subjacency, and this violation provides an account for its ungrammaticality. The second point to be mentioned concerns the trace of Wh-movement. Based on an argument of Rizzi (1980), Chomsky suggests that such a trace is not an anaphor for Opacity, not just in Italian, but in English, as well. The locality constraints on Wh-movement given by Subjacency instead for the relevant cases. This now partially unifies the treatment of Wh-trace, given the argument from strong crossover, discussed above, that Wh-trace has the behavior of a name. Interestingly, though, Chomsky explicitly limits the suggestion that Wh-trace is not an anaphor to Opacity effects, and not to NIC effects. Thus, he proposes that the difference between the mildly ungrammatical Subjacency violation (49) and the completely impossible (50) is to be attributed to the fact that NIC as well as Subjacency is violated in the latter.

(49) ?What, (s did you wonder (s who (s, t say, t, s)])

(50) *Who (s did you wonder (s what (s, t say, t, s)])

Taraldsen (1978), Pesetsky (1981), and Kayne (1980) presented analyses along similar lines. Chapter 5 [Freidin and Lasnik (1981)] argues against any such treatment, based on strong crossover phenomena. It is true that the standard examples of crossover, such as those presented in Chomsky (1976), all involved object NPs (hence, were not inconsistent with the proposal that a nominative Wh-trace is an anaphor). However, alongside the strong crossover examples presented by Chomsky, such as (35b) above, there are also examples where the trace is nominative. And contrary to what might be expected, such examples are no better than (35b) on the relevant interpretation:

(51) Who did he say t kissed Mary

Thus, to the extent that crossover provides a diagnostic, Wh-trace does not behave like an anaphor in NIC contexts any more than in Opacity.
contexts. Because of this, Chomsky (1981) rejects the NIC account of such an example as (50). To deal with this 'residue of the NIC' (RESNIC), Chomsky proposes a constraint specific to the distribution of traces, the E(npty) C(category) P(inciple), which will not be discussed here, but which is examined in detail in Lasnik and Saito (1984) [reprinted in Lasnik (1989)], Kayne (1984), and Chomsky (1986). See also Lasnik and Uriagereka (1988) and van Riensdijk and Williams (1986) and references cited in those books.

The technical details of Chomsky (1980) must now be considered, particularly with respect to the question thus far left open: How are disjoint reference effects to be handled? We begin with the theory of index assignment. First, movement involves obligatory coindexing of the moved category and its trace. Then, the remaining NPs in a sentence are indexed in 'top to bottom' fashion. '... an index is assigned to NP only when all NPs that e-command it or dominate it have been indexed.' For anaphors, Chomsky assumes that indices are assigned by 'rules of construal,' guaranteeing coindexation with an antecedent. As we will see, this introduces substantial redundancy into the system, since anaphors that are not bound in the appropriate local domain will be excluded regardless. To the extent that all relevant phenomena can be described in these terms, we will have a strong argument for free indexing.

The indices considered thus far, Chomsky calls 'referential indices'. A referential index is an integer. Nonanaphors are also assigned referential indices, though not, of course, by the same algorithm as in the case of anaphors. As before, the algorithm operates top down. But this time, coindexing, rather than coindexing, is what is forced. Each nonanaphoric NP that has not already received an index (by movement) is assigned an unused integer as its referential index. (In this case, as well, the stipulation that the index be a new one is ultimately unnecessary.) In addition to its referential index (an integer), each nonanaphor is also assigned what Chomsky calls an 'anaphoric index'. The anaphoric index of an NP is the set consisting of every integer that is the referential index of any NP c-commanding the NP in question. The (complete) index of a nonanaphor will be a pair \((r, A)\) where \(r\) is the referential index and \(A\) the anaphoric index. Given all of this, we will have such representations as (52):

\[
(52) \quad \text{John}_{[2,1]} \text{ told Bill}_{[3,2]} \text{ about him}_{[4,2,3]}.
\]

It is the anaphoric index of an NP that is responsible for disjoint reference effects. This is instantiated in the following way.

We will interpret the anaphoric index \(A = \{a_i, \ldots, a_n\}\) of \(\alpha\) to mean that \(\alpha\) is disjoint in reference from each NP with referential index \(a_j\). Thus, \(\text{him}_{[2,1]}\) in (52) is disjoint in reference from \(\text{John}\) and \(\text{Bill}\), and if \(\text{John}\) were to replace \(\text{him}\) (or \(\text{Bill}\)) in (52), the two occurrences of \(\text{John}\) would be disjoint in reference.

Notice that these indexing principles combined with this interpretive principle give the effects of both RI of Chomsky (1973) and the disjoint reference rule of Chapter 4 [Lasnik (1976)]. There is a conceptual difference, however. While the earlier analyses combined the syntactic and semantic aspects of disjoint reference into one rule, the proposal now under discussion separates the two aspects. There are syntactic rules and well-formedness conditions determining what syntactic representations are allowed. Then there is an interpretive principle assigning a meaning to a syntactically well-formed structure. In this clear separation of the syntax and semantics of anaphora, the analysis directly foreshadows that of Chomsky (1981), to be discussed below.

The present analysis in not yet complete. It does correctly rule out all cases of overlap in reference discussed in Chapter 4 [Lasnik (1976)]. Further, it rules out all cases covered by RI. But it does not yet capture the fact that RI showed TSC and SSC effects. Thus, it incorrectly excludes a coreferential interpretation for \(\text{John}\) and \(\text{he or him}\) in (53):

\[
(53) \quad \begin{align*}
\text{a. John thinks he is clever} \\
\text{b. John thinks Mary likes him}
\end{align*}
\]

Chomsky proposes to capture this property and the corresponding property of anaphor binding simultaneously. Specifically, he first proposes that the binding conditions (now NIC and Opacity) can be construed "as deleting certain indices from the anaphoric index of a pronoun, thus in effect blocking certain cases of disjoint reference and permitting reference to be free." For an NP with no anaphoric index (i.e., an anaphor) the binding conditions operate on the referential index, altering it in a certain way rather similar to the way in which the anaphoric index is altered in the case of a pronoun. The binding conditions will thus affect the 'designated index' of an NP, where the designated index of a pronoun is its anaphoric index, and the designated index of an anaphor is its referential index (which is, of course, its only index). These notions are made precise as follows:

\[
(54) \quad \text{Suppose that } \alpha \text{ has the designated index } j \text{ and } i \text{ is an integer such that } l = j \text{ or } i \in j. \text{ Then } \alpha \text{ is free}(i) \text{ in } \beta \text{ if there is no } \gamma \text{ in } \beta \text{ with index } i \text{ that c-commands } \alpha.
\]

\[
(55) \quad \text{Suppose that } \alpha \text{ has the designated index } j \text{ and is free}(i) \text{ in } \beta \text{ (} \beta = \text{NP or } S') \text{ where (a) } \alpha \text{ is nominative or (b) } \alpha \text{ in the domain of the subject of } \beta, \beta \text{ minimal.}
\]

\[
\text{Then } j \rightarrow 0 \text{ if } j \text{ is an integer, and } j \rightarrow (j - |i|) \text{ if } j \text{ is a set.}
\]

\(55a\) is the NIC and \(55b\) is Opacity. Consider the effects of (55) on the examples in (53). They will be initially indexed as in (56):
(56) a. John\(_{2\{1\}}\) thinks he\(_{2\{2\}}\) is clever
   b. John\(_{2\{1\}}\) thinks Mary\(_{2\{1\}}\) likes him\(_{2\{2\}}\)

Then the NIC will cause the 2 to be eliminated from the anaphoric index of he in (56a), and Opacity will have the same effect on him in (56b):

(57) a. John\(_{2\{1\}}\) thinks he\(_{2\{1\}}\) is clever
   b. John\(_{2\{1\}}\) thinks Bill\(_{2\{2\}}\) likes him\(_{2\{3\}}\)

The interpretive results will now be as follows: In (57a), neither John nor he is marked as necessarily disjoint from any other NP, hence these two NPs are free to corefer; in (57b), him is not necessarily disjoint from John, but him is disjoint from Bill and Bill is disjoint from John. Crucially, index subtraction never applies in the case of full lexical NPs: SSC and NIC are irrelevant to disjoint reference in these instances, as noted in Chapter 4 [Lasnik (1976)].

Anaphors are treated in terms of roughly the same formalism as pronouns, as stated in (55). An example such as (58) will have as its initial indexing something like (59):

(58) *John thinks Mary likes himself
(59) John\(_{2\{1\}}\) thinks Mary\(_{2\{1\}}\) likes himself\(_{2\{2\}}\)

In (59), himself is free(2) in the domain of a subject (Mary) in the embedded S’. That is, in that domain, there is no NP with index 2 that c-commands himself. Thus, its index becomes 2 = 2 = 0. To rule out the example, then, “it remains only to add that NP0 is not permitted in LF, where 0 is the referential index.”

Given this machinery, the mechanism specifically coindexing an anaphor with its antecedent becomes superfluous, as suggested above. Consider again (58), but now suppose that himself had been freely assigned an index, say 5:

(60) John\(_{2\{1\}}\) thinks Mary\(_{2\{1\}}\) likes himself\(_{2\{5\}}\)

Now himself is free(5) in the domain of a subject in the embedded S’. Hence, just as before, the index will become 0 (= 5 - 5), again running afoul of the LF prohibition. Simple examples behave in the same way. In (61), if himself is assigned any index other than the index of John, Opacity will change the index to 0, thus ruling out the representation.

(61) John likes himself
Further, even such an example as (62), where there is no potential antecedent, will be excluded in the same fashion.

(62) *Himself left

No matter what index is assigned to Himself, NIC will change that index to 0. Thus, there is no need for an initial indexing algorithm for anaphors. Indices can be freely assigned to them, incorrect results all being filtered out by the interaction of NIC, Opacity, and the LF filter. All that is needed is to specify an interpretation for coindexation, in particular that two NPs with the same referential index are coreferential. See Chapter 5 [Freidin and Lasnik (1981)] for the first presentation of this argument; and see below, and Chapter 6 [Lasnik (1981)] for further discussion of this issue.

Note, incidentally, that it is now obvious why an anaphor does not have an anaphoric index. Suppose that this were not the case. Then an example such as (61) would have an initial indexing such as (63):

(63) John\(_{2\{1\}}\) likes himself\(_{2\{1\}}\)

Since neither NIC nor Opacity is relevant here, (63) remains as the final representation. But the interpretive principle will now incorrectly demand that John and himself are disjoint in reference, a property that is not only not necessary here, but that is not even possible.

As with the formulation in (42) above, a certain class of long distance picture NP anaphors are permitted. Consider (44), repeated here as (64).

(64) They\(_{2\{1\}}\) expected that [pictures of each other,] would be on sale

In (64), each other is not nominative, nor is it free(2) in the domain of a subject. It is coindexed with the nearest c-commanding subject, They. Thus, neither NIC nor Opacity will apply, and each other will be allowed to keep its index. However, the potential problem noted earlier does in fact obtain. A coreferential pronoun in the position of each other in (64) is grammatical, yet it should not be. Consider (65).

(65) They\(_{2\{1\}}\) expected that [pictures of them\(_{2\{1\}}\)] would be on sale

Once again, neither NIC nor Opacity operates, hence the designated index of the pronoun them is unaffected. But now the interpretive principle demands (apparently incorrectly) that They and them be disjoint in reference, hence, non-coreferential. Thus, the grammaticality of (64) provides but little support for this approach. After all, any approach based on complementarity between anaphors and pronouns will get at best, and at worst, half of these cases right.

Chomsky (1981), henceforth LGB, presents a second major revision of Binding Theory, motivated by what he terms a number of conceptual and empirical difficulties with the approach in Chomsky (1980), henceforth OB. One of the difficulties, I have already discussed — that NIC incorrectly included Wh-trace in its domain. Recall that this is problematic
because, as indicated by crossover phenomena, Wh-trace has the behavior of a name rather than an anaphor. Among the further concerns that are addressed in Chomsky LGB, there are several more that are directly relevant to the present discussion. First, the properties of PRO were not all captured in the OB framework. In particular, the theory was too weak in not entailing that PRO occurs only in ungoverned positions (subject of infinitival or gerund), and too strong in excluding as SSC/Opacity violations certain cases of long distance control, as in (66).

(66) They thought I said that [PRO to feed/feeding each other] would be difficult

In (66), I is a subject establishing an opaque domain in which an anaphor must be bound, but the antecedent of PRO, They, is outside of that domain. Chomsky argues that there are a number of other problems with OB that are more conceptual in nature. I turn now to some of those.

Chomsky points out a sort of ‘redundancy’ between the theories of Case and of binding. In effect, both theories distinguish subject of infinitives as special: this is the one basic NP position that is not (normally) marked for Case; and it is also the one position generally transparent with respect to anaphoric phenomena. Chomsky raises the question whether Opacity “cannot somehow be reduced to Case theory just as the [TSC] was reformulated in terms of considerations of Case within the OB-framework, as the NIC.” A related question is what the two domains relevant for Binding Theory, subject of a tensed sentence and c-command domain of a subject, have in common. In terms of OB, these are in no way related.

Chomsky further suggests that the OB indexing conventions are another area where improvement is desirable. About these conventions, which were outlined in some detail above, Chomsky states,

While they work quite neatly, they are fairly complicated and it is worth asking whether it is not possible to eliminate the concept of “anaphoric index” entirely in terms of some more basic and simple notion.

It has already been noted that the convention for assigning referential indices is eliminable. Chomsky arrives at this conclusion in the course of his discussion. Momentarily, the issue of anaphoric indices will be considered. Chomsky further suggests that the phenomenon of disjoint reference itself is strange, the fact that pronouns enter into disjoint reference under the conditions where anaphors enter into coreference constituting “an odd state of affairs”:

Why should languages have this peculiar design, which in fact gives rise to the complexity of the indexing conventions and of the notion ‘free(i)’ defined in OB? Why shouldn’t pronouns have coreference, rather than disjoint reference, where, for example, reciprocals do?

To the extent that this state of affairs is problematic (and it is not entirely clear that it is), the problem is not really solved in the ‘GB’ framework of Chomsky (1981). Though the phenomenon is described in somewhat different terms, as we will see, it remains a basic unanalyzed property of pronouns that they show disjoint reference.

With these concerns as background, the central details of the GB approach to anaphora will now be presented. The indexing system assumed, contrary to the rather complicated one in OB, involves merely a single integer as the index of an NP. ‘Bound’ and ‘free’ are defined as in (67) and (68):

(67) \( \alpha \) is bound by \( \beta \) if and only if \( \alpha \) and \( \beta \) are coindexed and \( \beta \) c-commands \( \alpha \).

(68) \( \alpha \) is free if and only if it is not bound.

The domain in which an anaphor must be bound and a pronominal free is the ‘governing category’ of the item in question. (69) is the first approximation of the characterization of this central notion:

(69) \( \alpha \) is the governing category (GC) for \( \beta \) if and only if \( \alpha \) is the minimal category containing \( \beta \) and a governor of \( \beta \), where \( \alpha = NP \) or S.

The concept of government relevant here is just the one relevant to Case theory, Chomsky indicates. Thus, there is a sort of unification in the theory as a whole, though no explanation is attempted for why the same structural relation should be relevant to both modules. As in OB, three basic types of nominals are assumed: anaphors, pronominals, and fully lexical NPs (here called ‘R-expressions’). Recall that in OB, the differences among the three emerged in the mode of indexing, or reindexing. Anaphors received no anaphoric index, hence the designated index was the referential index. Pronominals and R-expressions had both anaphoric and referential indices, the former serving as designated index for a pronoun. Finally, R-expressions had no designated index, hence were not subject to reindexing at all. This three-way division now emerges in the form of the following three binding principles:

(70) A An anaphor is bound in its GC
    B A pronominal is free in its GC
    C An R-expression is free

All of these principles involve A-binding, that is, binding by an NP in a potential argument position. By A, anaphors are correctly excluded from nominative subject position, assuming that nominative Case, like other Cases, is assigned under government (by the agreement element AGR, presumably). Since no NP in an S c-commands the subject, a nominative anaphor will always be free in the minimal S containing it. This accounts for the ungrammaticality of the examples in (71), with himself and each
other the anaphors in (a) and (b), and the trace of NP movement the anaphor in (c).

(71) a. *John, believes that himself, is clever
    b. *They, believe that each other, are clever
    c. *Mary, is believed ti, is clever

Note that all of these are allowed with an infinitival complement:

(72) a. John, believes [himself, to be clever]
    b. They, believe [each other, to be clever]
    c. Mary, is believed [ti, to be clever]

(72a, b) are straightforward. In both cases, the subject of the complement is not governed within that complement. Either, each anaphor is governed by the exceptional Case marker believe in the higher clause. Thus, the matrix S is the GC in each case, and the anaphor is bound by the subject of the matrix. Similarly, in (72c), there is no governor of ti in the complement, hence the complement is not the GC for ti. At this point, there are actually two possibilities for ti in this configuration: either ti is governed by believe making the matrix the GC; or ti is not governed at all, with the result that it has no GC. While both possibilities correctly allow (72c), shortly it will become clear that the first is the correct one.

Given (70b) and (70a), pronouns, as usual, are predicted to be free precisely where anaphors are bound. This is illustrated in (73) vs. (74):

(73) *John, believes [him, to be clever]
(74) John, believes that he, is clever

Thus, the effects of the NIC for both anaphors and pronouns follow directly from Principles A and B of the GB binding theory. The core SSC cases follow as well. Consider (75) and (76):

(75) *John, believes [Mary to like himself,]
(76) John, believes [Mary to like him,]

In these examples, the pronoun and anaphor are governed by the verb of the complement clause like. Hence, the GC for each is the complement clause. (75) is correctly excluded by Principle A, since himself is not bound in its GC. And (76) is correctly allowed by Principle B since him is free in its GC. Finally, Principle C gives the same distribution for R-expressions as was given by the OB assignment of anaphoric indices. The fact that locality is not relevant for R-expressions was captured by the lack of a reindexing algorithm for R-expressions, and is now captured by the irrelevance of governing category to Principle C. The examples in (77) are thus correctly excluded. While John is free in its GC, it does not satisfy the much stricter requirement of being entirely free:

(77) a. *He, believes that John, is clever
    b. *He, believes Mary to like John,

With respect to the cases of anaphors and pronouns considered thus far, one of the problems Chomsky pointed out for OB has apparently been eliminated. It is now evident what nominative NP and NP in the domain of a subject have in common: both are governed internal to the minimal S dominating them, as seen in (73)—(76) above. Hence, both have that minimal S as a GC. Significantly, this result obtains with no stipulation. Seemingly, Chomsky has been successful in reducing two conditions to one. However, not all of the relevant phenomena have yet been considered in this new light. Recall that SSC effects show up not just in clauses but in NPs as well. For example, Chomsky (1981) presents the following contrast:

(78) a. *We, heard [their stories about each other,]
    b. We, heard [some stories about each other,]

In both (78a) and (78b), each other is governed by about. By (69), then, the bracketed NP is the GC for the anaphor. This correctly disallows (78a), but (78b) is incorrectly excluded as well. This is precisely the type of contrast captured in earlier work by the SSC, but it is not yet captured by the GB reformulation. Consequently, (69) is replaced by (79):

(79) \( \beta \) is a governing category for \( \alpha \) if and only if \( \beta \) is the minimal category containing \( \alpha \), a governor of \( \alpha \), and a SUBJECT accessible to \( \alpha \).

Chomsky proposes that AGR is the SUBJECT in a finite clause; standard subject (NP of S) is the SUBJECT of an infinitival clause; and NP of NP (if present) is SUBJECT of an NP. Note first that this correctly distinguishes between (78a) and (78b). In the latter, the object NP has no SUBJECT, and the GC is thus the matrix clause, the desired result. In (78a), on the other hand, the NP does not have a SUBJECT, hence constitutes a GC for the anaphor, with the result that the anaphor is, as desired, in violation of Principle A. Conversely, in these environments a bound pronoun will correctly be allowed. For clausal SSC cases such as (75) and (76), the change has no important consequence for the determination of GC, and the correct results are still obtained. Finally, the results are the same as before for NIC examples like (72)—(73); in these cases AGR now serves as both the governor for the nominative subject and as the relevant SUBJECT.

While (78) is now correctly handled, given the notion SUBJECT, it is somewhat less clear than before that the conceptual problem that concerned Chomsky has been entirely eliminated. The initial formulation of GC in (69) truly did reduce two domains (nominative NP, as in the NIC, and e-command domain of a subject, as in the SSC) to one. But for the
modification in (79), which was demanded by the facts in (78), much of the reductive work is done by the definition of SUBJECT. And the two parts of that definition are strikingly reminiscent of the original two domains: AGR establishes a GC for a nominative subject, and subject establishes a GC for any NP in its domain. Thus, the reduction is only successful to the extent that INFL containing AGR, and subject constitute a natural class. Chomsky suggests that they do:

The notion SUBJECT accords with the idea that the subject is the "most prominent nominal element" in some sense, taking INFL to be the head of S.

This is not implausible, but it is not clear that this rationale provides the correct division of cases. It is the head of S (rather than the subject of S) that counts as the most prominent nominal in that domain, hence, the SUBJECT. But it is crucially the subject of NP, rather than the head N of NP, that counts as SUBJECT in that case. If this were not so, both (78a) and (78b) would be excluded by Principle A, since, patently, both bracketed NPs have nominal heads. Thus, in contradistinction to the situation with Ss, presence or absence of subject is key here.

One final detail of (79), the mention of 'accessibility', remains to be discussed. Recall that certain cases of long distance anaphors have been problematic for most versions of BT. Example (64), repeated here as (80), is representative:

(80) They, expected [that [[pictures of each other],]] would be on sale

The embedded clause (and the subject of that clause) must not be a GC for each other or Principle A will incorrectly be violated. The picture NP has no SUBJECT, so it is immediately excluded from consideration. The complement clause does, however, have a SUBJECT — AGR. Accessibility is defined precisely in such a way that AGR will not be an accessible SUBJECT for each other:

(81) \( \alpha \) is accessible to \( \beta \) if and only if \( \beta \) is in the c-command domain of \( \alpha \) and assignment to \( \beta \) of the index of \( \alpha \) would not violate (82).

(82) \( [[y \ldots \delta \ldots]] \), where \( y \) and \( \delta \) bear the same index.

(82) is often referred to as the 'i-within-i condition'. Its incorporation into (81) guarantees that a SUBJECT will not be accessible to a given anaphor if that SUBJECT is (in fact) coindexed with a category properly containing the anaphor. Given that the subject of a finite clause is always coindexed (via superscripts in LGB) with AGR, this being the instantiation of subject-verb agreement, that AGR will never be accessible to anything within that subject. Consider (80) once again, but with AGR made explicit:

(83) They, expected [that [[pictures of each other], \{AGR\} would be on sale]]

Here if we were to assign to each other the index of AGR, 2, an i-within-i configuration would be created, since the NP pictures of each other has an index 2. Thus, the embedded clause is not the GC for each other, and the latter is free to seek an antecedent in the matrix. Note incidentally, that AGR in the matrix clause will be accessible to each other, since that AGR is not coindexed with any category properly containing each other. The consequence of this observation is that the matrix is the GC in this case. In the present example, this is a matter of little import. But if a structure like (83) is further embedded, the GC will be correctly determined to be that embedded structure itself. As (84) shows, each other is not free to seek its antecedent in a larger domain:

(84) \*We, think [that they, AGR \{AGR\} expected [that [[pictures of each other], \{AGR\} would be on sale]]]

Simple NIC configurations will still be ruled out, as the i-within-i exemption will not be applicable. Consider (85):

(85) \*John, thinks that himself, AGR \{AGR\} will win

Assignment of the index of AGR to himself would not result in an i-within-i configuration. Hence, AGR is accessible to himself. Since AGR also governs himself, the GC for the latter is the embedded clause. Since it is not bound in that domain, Principle A is violated.

It might be noted that one minor problem noted above in connection with the OB approach is now eliminated. Recall that a small residue of the redundancy that Chomsky claimed to have eliminated still remained. While the change from TSC and SSC to NIC and Opacity removed the overlap in conditions in a simple example like (40), repeated as (87), (43), repeated as (88) was still ruled out by two binding conditions:

(87) \*They told me [\{I, what I gave each other\]

(88) \*The men think that Mary said [that each other would win]

In (88), according to OB, each other is both free in the domain of a subject (Mary), and is a nominative anaphor free in S'. However, given the GB revision, (88) is excluded in precisely one way. Given the fact that an anaphor has (at most) one GC, there is no redundancy. The most deeply embedded clause in (88) is the GC for each other, and each other is free in that GC, in violation of Principle A.
Another somewhat more substantial problem still remains. Like OB, GB predicts essentially total complementarity between anaphors and pronouns. In the latter instance, this follows from the fact, stated in (70), that an anaphor must be bound and a pronoun free in precisely the same domain — the GC. But, as before, the device (in this case, accessibility) that allows long distance anaphor binding in limited circumstances, as in (83), will necessarily (and incorrectly) exclude a bound pronoun in the same circumstances:

(89) They, expected [that [[pictures of them],\textsuperscript{2} AGR\textsuperscript{2} would be on sale]]

The GC for them in (89) is the matrix clause, and it is bound, hence not free, in that clause, in violation of Principle B. The problem of lack of complete complementarity is, thus, still not solved, or, indeed, even addressed. It is not until Chomsky (1986), to be discussed below, that Chomsky offers an account of this phenomenon.

A major innovation in the GB approach to anaphora involves the treatment of PRO, the base-generated null subject of non-finite clauses, as in (90):

(90) John tried [PRO to leave]

As noted above, Chomsky argues that a defect of OB was that it failed to capture the central distributional fact that PRO must be ungovernmented. Thus, the examples in (91) are all ungrammatical regardless of the index on PRO:

(91) a. *I like PRO
b. *John believes PRO to be intelligent
c. *Mary believes PRO is intelligent

The analysis of this fact offered in LGB is as follows. PRO is similar to overt pronouns in that it may have an antecedent, but not within its own clause or NP. On the other hand, PRO resembles anaphors in lacking intrinsic referential content. Rather, it is either assigned reference by an antecedent (the phenomenon of Control) or is indefinite in interpretation (so-called arbitrary PRO). Suppose, then, that PRO is a pronominal anaphor. By virtue of this, PRO is subject to both Principles A and B. If PRO has a GC, it must then be both bound and free in that GC, a contradiction, since free means not bound. Thus, PRO has no GC, hence, is ungovernmented. The central property is therefore derived.\textsuperscript{4}

This elegant deduction of the central distributional property of PRO clearly relies on the fact that the definition of GC makes reference to government. Is there independent justification for this aspect of the definition? Chomsky considers a simplification whereby GC (79) is changed to "binding category" as follows:

(92) \( \beta \) is a binding category for \( \alpha \) if and only if \( \beta \) is the minimal category containing \( \alpha \) and a SUBJECT accessible to \( \alpha \)

Principles A and B would then be stated in terms of (92):

(92) A An anaphor is bound in its binding category
B A pronominal is free in its binding category

Chomsky observes that this modification has "no effects for elements that are governed since for such elements the governor will always be contained in the binding category." For PRO, however, Chomsky shows that the change leads to an incorrect consequence. Consider first (93), where the modified theory still works correctly:

(93) *John, expected [him, to win]

Here the GC for him is the entire sentence, since in the embedded clause there is no SUBJECT accessible to him. (93) is thus correctly excluded as a violation of (92B). The problem is that by exactly the same line of reasoning, (94) will be incorrectly excluded.

(94) John, tried [PRO, to win]

Just as in (93), the matrix is the binding category, since it contains a SUBJECT, John, accessible to PRO (and no smaller domain contains such a SUBJECT). Consequently, (94), like (93), violates (92B). Reference to government, as in (79) but not (92), is apparently necessary to correctly distinguish (93) from (94): PRO in (94) is not governed, as an S' boundary, a barrier to government, intervenes between it and tried. Thus, there is at least a bit of independent evidence that the correct characterization of the domain relevant to binding theory involves government.

As we have seen, if an item is ungovernmented, then it has no GC. But there is apparently an additional way for an item to have no GC, namely, for it to have no accessible SUBJECT. This additional possibility has undesirable consequences. Consider (95):

(95) *For each other to win would be unfortunate

In this example, though each other is governed (by for), it nonetheless has no GC according to (79), as there is no SUBJECT accessible to it. (95) then incorrectly fails to violate Principle A. One might imagine that (95) simply lacks any possible interpretation, quite apart from any considerations of binding theory. However, Chomsky rejects such an account of (95), observing that such an example as (96) could not be excluded on such grounds.

(96) *For each other, to win would be unfortunate for them,

Chomsky argues that the simplest rule of interpretation for an anaphor
would apply to any coindexed pair of NPs including the anaphor. But this simplest rule would, in fact, assign an interpretation to (96). Chomsky concludes that (96) must then fall under Principle A:

The point is that the structural relation between the antecedent and the anaphor should be expressed in the binding theory, not by a (quite generally redundant) stipulation added to the rule of interpretation.

Consequently, (97) is added to the theory:

(97) A root sentence is a governing category for a governed element

In both (95) and (96), each other is governed. Hence, the root sentence is a GC in each case, and the anaphor is not bound in that domain (in fact, is not bound at all). Both examples now violate Principle A, as desired. The identical account is available for (98), and even (99):

(98) *Pictures of himself are on sale

(99) *Pictures of PRO are on sale

In both examples, there is a governed anaphor, himself in (98) and PRO in (99). The root S is thus a GC, and the anaphor is free in that GC, violating Principle A, the desired result. Note, by the way, that the reference to government in (97) is crucial. To see this, consider an alternative such as (100):

(100) A root sentence is a governing category for an element that otherwise has no governing category.

For the examples examined so far, it can be easily seen that (100) gives just the same results as (97). However, it has the completely unacceptable consequence that PRO will never be possible. This is so because given (100), PRO, like any other item, would invariably have a GC, even if it is not governed. But once it has a GC, the standard contradiction arises: it cannot be both bound and free in the same domain, hence either Principle A or Principle B would necessarily be violated. Thus, the required extension of the definition of GC must be (97), which grants the crucial exemption to an ungoverned item, rather than (100), which does not.

Chomsky points out a substantial conceptual change resulting from some of the essentially technical innovations discussed thus far with respect to LGB. Recall that in the OB approach, while disjoint reference was explicitly displayed in a representation (via the device of anaphoric indices), coreference between non-anaphors was not. That is, when a pair (R-expression, pronoun), say, were not marked as disjoint, then they were free to corefer or not. This was the formal instantiation of the free reference for pronouns of Chapter 4 [Lasnik (1976)]. Thus, as discussed in Chapter 6 [Lasnik (1981)], the OB theory made provision for three possible referential relations among NPs: disjoint reference (indicated by anaphoric index); bound anaphora (indicated by identical referential indices); and free reference. Since in LGB, every NP has exactly one index (the referential index of OB), only two possible relations between NPs are expressible. That is, two integers are either the same or different. Concerning these two possibilities, Chomsky proposes the following:

... pronouns are "proximate" if they are coindexed with some other element and "obviative" if not coindexed with any other element ... This is ... a departure from the spirit of Lasnik's well-known proposal (Lasnik (1976)) that pronouns are free in reference, subject to other conditions.

As remarked earlier, the syntax of the LGB system is simpler than that of OB. But as also suggested, it appears to be too simple. That is, distinctions that evidently need to be made cannot. Chapter 6 [Lasnik (1981)] discusses this problem in detail, arguing that there are at least three linguistically significant referential relations among NPs: coreference, disjoint reference, and overlap in reference. Since the syntax provides but a two-way distinction, the theory is inadequate. It is for this reason that Chomsky (1981) concedes the analysis that he develops that "The proposal does not appear to be feasible ..." Chomsky further suggests that this problem might be solved "by extending the theory of indexing to use of more complex indices ..." Chapter 9 [Lasnik (in press)] considers this possibility by exploring a theory in which a referential index is not a single integer but is, instead, a set of integers.

Chomsky (1982), henceforth C&C, adopts the LGB approach to binding, along with certain minor elaborations, modifications and extensions. First, an explicit feature representation of nominal types is introduced: NPs are combinations of [+/-a(naphoric), +/-p(ronominal)], with [+a] subject to Principle A and [+p] subject to Principle B. Pure anaphors, lexical ones and NP-trace, are [+a, -p]. Pure pronominals are the reverse, [-a, +p]. It is in C&C that the null analog of lexical pronouns, 'pro', is introduced. pro is argued to fill a gap in the paradigm of null NP types, and to provide a principled empty category to serve as subject in null subject languages. PRO is [+a, +p], hence subject to both Principles A and B. The PRO theorem thus holds precisely as in LGB. There is no lexical [+a, +p], since such a category would have to be both bound and free in the same domain if governed, but would be illicitly Caseless if un governed. Also as before, lexical R-expressions and Wh trace constitute a sort of default, having the negative value for both features — [-a, -p].

While Principles A and B are carried over unchanged from LGB, Chomsky argues that Principle C is eliminable. This argument relies on a particular approach to empty categories that Chomsky puts forward in C&C. According to this approach, there is but one empty category whose varied behavior is determined by the varied contexts in which it finds
Chomsky (1986), henceforth K of L, presents a number of rather substantial modifications in the theory of binding. The first to be discussed here, initially suggested in C&C, concerns Principle C. Chomsky (1977) had argued that so-called tough-Movement constructions, as in (104), involve Wh-movement of an operator.

(104) John is tough [O [PRO to please t]]

Chomsky observes in C&C that the usual formulation of Principle C would have the incorrect effect of excluding such a representation, since John is presumably coindexed with t, resulting in the A-binding of the latter by the former. Noting that the O operator is also coindexed with the trace (via movement), Chomsky suggests that Principle C should be modified so as to preclude only local A-binding. In (104), though t is A-bound, it is not locally A-bound, because of the intervening A’-binder, O. The modification is instantiated as (105):

(105) An R-expression must be A-free in the domain of the operator that A’-binds it.

In K of L, this is restated as (106), which is to be understood as in phonology as two disjunctively ordered principles:

(106) An R-expression is A-free (in the domain of the head of its maximal chain).

In (104), the maximal chain of t is (O, t), and t is A-free in the domain of O, the head of the chain.

It is interesting to note that this is by no means the first time that problems involving tough constructions (and related ones) have necessitated certain complications in the theory. In fact, from the inception of modern Binding Theory in Chomsky (1973), these constructions have been problematic. Chomsky (1973) treated them by NP Movement, deriving (107) from (108).

(107) John is tough [PRO to please t]

(108) It is tough [PRO to please John]

Such an analysis captures the apparent relation between the two examples and, arguably correctly, makes the subject position a non-theta one. However, as Chomsky discussed, SSC would be violated by direct movement of the NP. Consequently, Chomsky proposed a two step derivation, with the first step involving PRO replacement:

(109) It is tough [John to please t]

This movement is entirely within the embedded clause, and the second movement, producing (107), is from an accessible position, subject of an
infinitive, just like standard raising. However, PRO replacement is quite a
dubious operation, as it moves an NP into an argument position. Further,
given the availability of such a derivation, it is difficult to see how to block
a similar one whose first step is a legitimate instance of NP movement:

(110) a. It is easy [r to be arrested John]
b. It is easy [John to be arrested r]
c. *John is easy [r to be arrested r]

Chapter 3 [Lasnik and Fieno (1974)] is an early attempt to reanalyze
these constructions in a way compatible with SSC but avoiding the
problems just outlined. Certain aspects of the analysis were incorporated
into that of Chomsky (1977), which, in its essentials, is still maintained in
Chomsky (1986).

Chapter 7 [Lasnik (1985)] discusses a binding problem potentially
involving the kind of redundancy considered on several occasions above.
As will become evident, the proposals in K of L bear on this redundancy.
The phenomenon at issue is of historical interest as well, since it is a rare
case where SSC and TSC markedly diverge from Principle A: the NP
movement in (111) would violate SSC and TSC, but, surprisingly, the
trace is bound in its GC.

(111) *John, is believed [(that) he, likes t1]]

An account of (111) in terms of Case, for example, via a prohibition
of Case-marked NP trace, results in massive redundancy for (112), which
would violate this Case constraint as well as the ECP and Principle A.11

(112) *John is believed [(that) r is intelligent]]

In part to eliminate a portion of this redundancy, Chomsky in effect
excises the NIC from the binding principles, leaving only the SSC as the
characterization of GC. As we will see, for Principle A, the revision
amounts to the requirement that an anaphor must select the nearest
potential binder.12 Principle A has always had something of this flavor,
except that AGR, which can never serve as an actual A-binder, nonetheless
blocked access to a higher binder in standard NIC contexts. If, as
proposed in K of L, AGR does not count as an accessible SUBJECT,
conceptual simplification results.

Before we proceed to the precise details of the K of L proposal, one
further phenomenon motivating the revision should be considered. As has
been observed earlier, while the theory of binding has been based on
complementarity between anaphors and bound pronouns, there are excep-
tions to this complementarity. (113) and (114) are two such configura-
tions.

(113) a. The children like [each other’s friends]
b. The children like [their friends]

(114) a. The children thought that [pictures of each other] were on sale]
b. The children thought that [pictures of them] were on sale]

As also noted earlier, and as discussed by Huang (1983), these paradigms
suggest that the GCs for anaphor and for pronouns are different. The
formulation about to be considered addresses both of the problems just
mentioned: redundancy, and lack of total complementarity. In the follow-

... the relevant governing category for an expression α is the least CFC containing a
governor of α in which α could satisfy the binding theory with some indexing (perhaps not
the actual indexing of the expression under investigation).

To the extent that the theory is able to state this, it will correctly allow
both (113a) and (113b). In the former, though the object NP is a CFC,
and though it has a governor for each other (since each other is assigned
Case), there is no indexing on which the anaphor could be bound within
that NP. Thus, the entire clause is the GC, and the anaphor is bound in
that domain. In (113b), on the other hand, the object NP is the GC for
the pronoun, since there is some indexing on which the pronoun is free within
that NP. In fact, it is free in that domain on any indexing. The examples of
(114) will also both be allowed. In (114a), there is no potential binder for
each other nearer than the matrix subject. Hence, the matrix clause is the
GC, and the anaphor is bound in that domain. In (114b), the NP pictures
of them seems not to be a CFC, since a grammatical function compatible
with the head pictures is not represented. If this is correct, then the fact
that them is free in that NP does not suffice. However, the complement
clause is clearly a CFC, and them is also free in that domain. Thus, there
is some indexing on which BT could be satisfied in that domain. That
clause is therefore a GC, and the pronoun is free in that GC. Note finally
that, as desired, Principle A is no longer applicable to (112). There is no
potential indexing on which t could be bound in the embedded clause,
hence that clause is not a GC. A portion of the redundancy noted thus
disappears: only ECP and Case requirements are relevant to (112).

This conception of BT does raise certain questions of its own. Before
proceeding to these, I present in (115) and (116) Chomsky’s formulation.
I omit Principle C, since it is unaffected by the considerations under
discussion. In (115) and (116), E is an expression with indexing I, α is an
NP, and β is a domain. The indexing I and the pair (α, β) are ‘BT-
compatible’ if α satisfies the binding theory in local domain β under
indexing I. (116) is to be understood as a "licensing condition ... for a category α governed by a lexical category γ in the expression E with indexing I ..."

(115) I is BT-compatible with (α, β) if:
(A) α is an anaphor and is bound in β under I
(B) α is a pronoun and is free in β under I

(116) For some β such that (a) is true, I is BT-compatible with
(α, β):
(a) α is an anaphor or pronoun and β is the least CFC containing γ for which there is an indexing J (not necessarily the actual indexing) BT-compatible with (α, β)

Since this theory is designed to allow environments in which either anaphors or bound pronouns can occur, the first question to ask is whether PRO is correctly excluded from such positions. That is, if a position is consistent with the requirements on anaphors, and independently consistent with the requirements on pronouns, why would it not be compatible with those of a pronoun anaphor, incorrectly allowing (117)?

(117) a. *The children like [PRO friends]
b. *The children thought that [[pictures (of) PRO] were on sale]

The answer is, in some respects, a rather unsatisfyingly technical one. Consider, for example, (117a) in the light of (115)–(116). The object NP is the least CFC containing PRO and a governor for PRO. Further, there is a potential indexing, namely any indexing, under which PRO is free in that domain, in evident satisfaction of the licensing condition for pronouns. Now note that if PRO is coindexed with The children, the licensing condition for anaphors will be satisfied as well. The matrix is the smallest CFC containing a governor for PRO in which there is a potential BT-compatible indexing. But, by hypothesis, this potential indexing is once again the actual indexing, and the requirement seems to be satisfied. The solution to this apparent problem lies in the precise statement of (116): "For some β" takes scope over the entire constraint. Thus, there must be precisely one domain in which the licensing requirements are satisfied. The analysis of (117) presented immediately above is thus excluded, since it relied on two different β's, the object NP and the matrix S. This now raises the question of how PRO is ever possible, even in an un gover ned position. The answer is that the above licensing condition is strictly for governed α. Interestingly, it is still true that explicit reference to government cannot be avoided.

Exploring the details of K of L further, note that the stipulation (97) in LGB, motivated by examples such as (98) and (99), repeated here as (118) and (119), is no longer needed.

(118) *Pictures of himself are on sale
(119) *Pictures of PRO are on sale

Recall that in the LGB theory, since himself and PRO in these examples had no accessible SUBJECT, they had no GC, an incorrect result. Consequently, it had to be stipulated that the root sentence is the GC in such cases. But now, the ungrammaticality of (118)–(119) follows directly. (116) is a licensing condition for a governed anaphor, but is not satisfied in either example, since in neither case is there a β with a possible indexing J BT-compatible with (α, β). That is, there is no NP that c-commands the anaphor. Thus, himself and PRO are not licensed. (120), on the other hand, will be allowed:

(120) Pictures of him are on sale

Here, the matrix is a CFC containing a governor for him, and there is a possible BT-compatible indexing J; in fact, any indexing is BT-compatible, since the requirement on a pronoun is that it be free.

Traditional SSC effects, as in (121), are easily accommodated:

(121) a. *John believes [Mary, to like himself,]
b. John believes [Mary, to like him,]

In (121a), licensing for himself fails. The embedded S is the least CFC containing a governor for himself in which himself could be bound, but it is not. In (121b), on the other hand, licensing succeeds. The embedded S is the least CFC containing a governor for him in which him could be free, and it is free in that domain.

Turning now to NIC effects, we have already seen that part of the motivation for the K of L theory is that it eliminates the redundancy evident in (112) between NIC and other conditions by, in effect, eliminating NIC. However, there were other traditional NIC effects involving no such redundancy, and these now are unexplained. Consider (122):

(122) *John believes [(that) himself, is intelligent]

There is no potential binder for himself in the embedded clause. Consequently, the matrix is the domain in which (115A) must be satisfied, and it is so satisfied. Thus, there is no BT violation in (122). Further, since John and himself do not constitute an A chain, the Case requirement alluded to in connection with (112) is not relevant. Finally, the third constraint relevant to (112), the ECP, appears not to be relevant here, because there is no trace. However, based on a proposal of Lebeaux (1983), Chomsky
argues that this is only an appearance. Chomsky suggests that at LF, lexical anaphors must undergo movement,

... rather in the manner of reflexivization in the Romance languages, with a reflexive clitic binding a trace ... English would have LF-movement of the anaphor corresponding to the S-structure representation in the Romance case ...

Then, if ECP is implicated in (112), it would also be relevant to (122). In both instances, there would be a trace in a position not susceptible to proper government. In a sense, in connecting conditions on traces and lexical anaphors in this way, the theory has come full circle. Recall that in OB, there was an attempt, abandoned in LGB in favor of the ECP, to derive a constraint on Wh-trace from the NIC. Now in K of L, we have the same connection, but in reverse: an analysis of NIC in terms of ECP. The resulting theory handles a substantial range of cases in a reasonably straightforward way, and the rather problematic notion accessible SUB-JECT is eliminated. With respect to the recurrent themes of this chapter, the near complementarity between anaphors and bound pronouns is accommodated in essentially the way that it always was, but now the residue of non-complementarity is handled as well. Further, virtually no redundancy remains internal to BT, and some of the redundancy between BT and other modules has also been eliminated.

Chomsky's successive modifications in BT have been summarized and illustrated in this chapter, but it is important not to overemphasize these changes. In important ways, the most recent theory discussed here, that of K of L, is still very closely related to the earliest, that of Chomsky (1973). Conceptually, the requirement that a pronominal must be free in a certain domain, which is now enforced by (115) and (116), is not drastically different from RI, illustrated in (5) - (7). Further, the central distinction between pronouns and anaphors, which was implicit in Chomsky (1973) and explicit in Chomsky (1976), is as important as ever. Even the notion of a domain defined (at least in part) by subject is in evidence early, the SCC (1a), and late, CFC. The extension of the insight behind RI, that the theory should incorporate a device disallowing coreference, from the local affects of RI to more distant ones in Lasnik (1976) remains in the form of Principle C. In fact, the developments explored here can best be seen not as a series of revolutionary upheavals in the study of anaphors, but rather as the successive refinement of one basic approach, and one that has proven remarkably resilient. Given that BT has become the subject of intensive investigation, with new phenomena in previously unexplored languages being constantly brought to bear, and all this while old problems from familiar languages remain, further refinements, or even revolutionary upheavals, are inevitable.

NOTES

1 Chomsky appears to assert exactly the opposite of this, following his example (54). But this is only an appearance. His phrase structure rules reversed the convention of Bresnan (1970), on which they were based, and they were also the reverse of the now standard convention. Thus, instead of $S \rightarrow \text{COMP} \, S$, he had $S \rightarrow \text{COMP} \, S^+$.

2 It is interesting to note that successive cyclic movement is thus generally enforced by both (i) and Subjacency, the other major constraint first proposed in Chomsky (1973). This sort of 'redundancy' (though not this particular case) was already a concern in Chomsky (1973). It was to become a recurrent issue.

3 Semantically, this would already be problematic, of course.

4 Note, however, that SSC must not be constrained by Subjacency. (i) is generally not regarded as significantly better than (ii):

(i) *They believe Mary likes each other

(ii) *They believe Mary to like pictures of each other

5 Later, this will be qualified, in light of the fact that the example is far worse than an ordinary 'island' violation.

6 The terminology is somewhat unfortunate, since anaphors crucially do not have anaphoric indices. Only nonanaphors do.

7 One obvious minor difference is that the rule in Chapter 4 is stated in terms of precedence and 'command', while subsequent work assumed the purely hierarchical c-command, introduced in Reinhart (1976). But see Chapter 8 [Barrs and Lasnik (1986)] for some evidence that linear order might, in fact, be relevant to anaphora.

8 Momentarily, we will see that the 'PRO theorem' does not follow in its entirety from these considerations. Only under stipulation, albeit one with some independent justification, will be needed. It might also be noted that the PRO theorem rather strongly relies on Principles A and B involving the same domain. If, for example, the domain in which an anaphor had to be bound were larger than the domain in which a pronoun had to be free, then a governed pronominal anaphor would, in principle, be able to exist. Thus, apart from the near total complementarity of overt anaphors and pronounals, the fact that governed PRO is always ungrammatical provides further evidence that the relevant domains for Principles A and B are, at least, very similar.

9 It is not entirely clear what the terms 'proximate' and 'obviative' in the following quotation mean. Indeed LGB there is a characterization given, but it cannot be what is intended here, for it would be largely circular: "... we will call a pronoun or PRO "proximate" when it is coindexed with an antecedent and "obviative" when it is not."

10 A category is locally A/A'-bound if its closest binder is in an A/A'-position. Note that (101) does not allow for a pure pronounal empty category. For this reason, (ii) is ultimately modified to allow optional assignment of [+a].

11 The theory of Case developed in K of L extends to (i), since it is proposed that even adjectives and nouns assign Case (what Chomsky calls 'inherent Case'):

(i) *John, is believed [that [he, is proud, i]]

However, even this view of Case does not extend to (ii), since, crucially, inherent Case is only assigned under theta marking, and belief in (ii) does not theta mark the position of t.

(ii) *John, seems [that [his, belief [i, to be intelligent]] is absurd]]

cf. It seems that John's belief that he is intelligent is absurd

See Lasnik and Uriagereka (1988) and Lasnik and Saito (in press) for further discussion.
12 Earlier, Huang (1983) suggested an approach to the characterization of GC that was conceptually rather similar to this one, motivated by considerations to be considered immediately below.

13 See Aoun (1985) for an approach to these phenomena conceptually similar to the one in OB, but foreshadowing certain aspects of the one in K of L.