Chapter 2:
Some Architectural Issues in a Minimalist Setting

Minimalism (at least as presented here) takes GB as its starting point. The reason for this is twofold. First, GB is a very successful theory of grammar with a very interesting theoretical structure and ample empirical coverage. The former property provides grist for the methodological concerns that minimalism highlights. The latter property permits discussion to move beyond mere methodology (which, as Chomsky has pointed out elsewhere, can be little more than a form of harassment) by setting empirical bars for prospective theories to clear. Second, GB is the most fully worked out version of a principles and parameters (P&P) approach to UG and, as noted in the introduction, all minimalist theorizing takes for granted a P&P architecture. As such, considering a GB style theory from the vantage point of minimalist methodological concerns is a good way of getting into substantive issues quickly. So, let's start with a quick listing of GB's major properties.

1. A Basic GB architecture: What are the main properties of a GB style theory?

A. First and foremost, it has a P&P architecture. This means that UG is composed of principles with open parameter values that are set by experience, I.e. By the PLD. The driving force behind P&P theories is the need to answer Plato's problem in the domain of language. By having innate general principles with open parameter values one can deal with two basic facts that characterize language acquisition: (a) it is very fast despite a very serious deficiency in the data that the child can use in fixing its competence (b) the surface variation that various languages display. This dual problem is easily accommodated if P&P is roughly correct. The ease of acquisition is due to the rich innate principles that the child comes equipped with. The variation can be traced to the fact that different parameter values can result in significantly different outputs.

B. GB theories identify four significant levels of the grammar: DS, SS, LF and PF. These levels are phrase markers with specific functional and substantive characteristics. Let's consider these.
     DS, is, substantively, the phrase marker at which "pure GF theta" is represented. This means, it is where an expression's logical/thematic role perfectly coincides with its grammatical role. Logical subjects are DS (grammatical) subjects, logical objects are DS (grammatical) objects etc. Positions with no thematic logical import are left empty. Positions that are thematically active are all filled. The thematic requirement in (1) can be taken as part of the definition of DS.

> (1) DS is the phrase marker at which all thematic positions are filled/saturated and no non-thematic positions are filled/saturated.

     An example or two will help fix ideas. Consider a predicate like 'persuade'. Thematically, it requires a "persuader", a "persuadee" and a propositional complement. In a sentence like (2a) this requires that the subject, object, and propositional theta roles be discharged at DS by having the subject position, object position and sentential complement position filled in DS. This means that there must be something in the subject position of the embedded clausal complement, though it is phonetically null, call it ec (empty category).

> (2) a. John persuaded Harry to leave
> b. John persuaded Harry [ec kiss Mary]

This ec is an obligatorily controlled PRO in GB whose antecedent is 'Harry'. The details of this construction are not important here. What is key is that given a notion of DS, this position must be filled by something because the embedded verb 'kiss' has a thematic role to discharge, the "kisser", and so the position must be filled by something. Here that something is PRO.

This contrasts with raising constructions in which positions are not filled. Consider (3a). The matrix subject position of appear is not a thematic position given that it can be filled by an expletive 'it' (3b). This means that 'John' in (3c) was not base generated in that position but must have gotten there transformationally. Thus, at DS, the subject matrix position is filled by nothing at all, not even a null expression. I designate this position with '0'.

(3) a. John appears to have kissed Mary
     b. It appears that John kissed Mary
     c. [0 appears [John to have kissed Mary]

DS can also be functionally characterized. In GB, it is the phrase marker that is output of phrase structure operations plus lexical insertion and it is the input to transformational operations. By being the locus of PS rules, DS is the locus of a grammar's generativity.

There is some interesting evidence for DS within GB. The best of it revolves around distinguishing raising from control which we will return to later on. There is also some interesting evidence against the existence of a DS level that we will review when we consider minimalist objections to DS.

SS can be substantively characterised as the point in a derivation in which the derivation splits, sending one copy off to PF for phonetic interpretation and one copy to LF for "semantic" interpretation. It is also the place where several grammatical modules ply their trade. Thus, case is assigned at SS, some aspects of the binding theory are applied there, 0-operators are identified there, some aspects of the ECP apply there (argument traces are gamma-marked at SS) and subjacency has been argued to apply there. In addition, SS is useful in distinguishing languages that are identical semantically but differ with respect to overt phrase structure. Thus, French differs from English in overtly raising V to T. In English, this V to T movement happens covertly at LF. Chinese differs from English in moving Wh's to CP at LF while English does this overtly.

It is fair to say, that SS is the queen of GB levels. It is the most theory internal level of the grammar. And, a large number of modules apply there to filter out unwanted derivations. One of the most interesting set of arguments spawned by the Minimalist program argue that SS is both dispensable and undesirable. We return to these below.

PF and LF are interface levels within GB. This means that they provide the grammatical information required to assign a phonetic and semantic interpretation to a sentence. Various proposals have been put forward about what operations apply at these levels. The most important of these is the ECP filter that functions to weed out derivations with unlicensed traces at LF. The binding theory and control are also thought to apply at LF. It is less clear that anything applies at PF, however, it has been argued that head-government is a PF process, viz. that applies on the PF side of the grammar.

C. The grammar has a T-type of organization. There is an important distinction between Movement operations that provide phonetically overt effects and those that do not. The latter apply on the LF side of the grammar and hence are phonetically invisible. The Grammar looks as in (4).

$$(4) \quad DS \xrightarrow{\text{Move}} SS \xrightarrow{\text{move}} \begin{array}{c} LF \\ \\ PF \end{array}$$
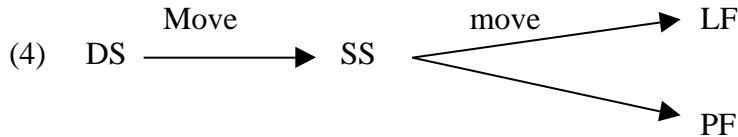
The Move operation that applies post SS on the LF side of things is the same operation that applies before the split. The only difference is that the outputs of this application of Move do not feed PF. Operations that have been postulated to so apply include QR, Associate to Expletive Raising, Anaphor movement, and WH-raising. Thus, there is a single movement operation that applies pre and post SS with different manifestations phonetically.

D. The projection principle applies to all levels of derivation. This makes derivations monotonic in that information from earlier structures, such as thematic information, must be preserved at all levels of derivation, in particular, DS,SS,LF (PF is not so constrained). One consequence of this is that traces are required to preserve the thematic and structural information coded in DS. Thus, the projection principle requires that if a V takes an object in DS then it must take one at SS and LF as well. This implies that if an expression moves, then some residue of its prior position must be maintained or the verb will "detransitivize", violating the projection principle. In effect, the projection principle forces all movement to leave traces behind to mark the positions from which movement occurs.

Within GB, the EPP is generally augmented to include a stipulation that all clauses must have subjects. This is the "extended" EPP. The principle arguments for this are empirical. We return to them in later chapters.

E. GB embodies a very simple transformational component. It includes two rules, Move and Bind. The latter allows free indexing of DPs. The former, allows anything to move anywhere anytime. Moving an XP leaves behind traces (due to the projection principle) which have the form '[XP e]', I.e. An XP category with null phonetic content. Traces are by definition silent and are by definition co-indexed with what has been moved. Traces are inherently problematic entities within GB in that they must be licensed. The ECP is the module of the grammar that concerns itself with licensing traces.

These two very general rules massively overgenerate unacceptable structures. To compensate for these very general rules GB grammars deploy a group of information specific modules.

F. GB consists of a host of interacting modules that act as filters to bar unwanted overgeneration and "prepare" a phrase marker for interpretation at LF and PF. These modules track Case features, Theta roles, Binding configurations, trace licensing (ECP and Subjacency), phrase structure (X'), and Control relations. These modules license these different kinds of information at different points in a derivation. Phrase markers that fail to conform to the required modular specifications are weeded out. This, serves to clean up the mess that application of the very general rules of Bind and Move generate. It also serves to ready phrase markers for interpretation e.g. Case theory tells one how to phonetically interpret a pronoun, binding theory accounts for which DP serves as antecedent for an anaphor etc.

G. The fundamental grammatical relation within GB is government. The conceptual unity of GB modules resides in their conditions exploiting the common relation of government. As noted, the kinds of information that GB modules track are very different. Theta roles are

different from case features, anaphors are different from bounding nodes, reflexives are not empty categories. What lends conceptual unity to these diverse modules is the fact that their reach/applicability is limited to domains defined in terms of government. Case is assigned under government, as are theta roles. Binding is checked within minimal domains that are defined using governors. The ECP (and Bounding Nodes) are stated in terms of Barriers which are in turn defined via government. There is thus an abstract conceptual unity provided by this key relation to otherwise very diverse modules.

There are various incarnations of GB. However, the overarching structure of these different implementations share most (if not all) of the general architectural features noted above. In the remainder of this chapter, we begin the task of reconsidering the status of these broad systemic features of GB against the methodological backdrop that minimalism suggests.

2. Minimalist Qualms

Despite its successes, there are reasons for rethinking the standard GB assumptions, at least from a minimalist point of view. Recall the picture that animates the minimalist enterprise: are the minimal boundary conditions on any adequate P&P theory also maximal? We fleshed these minimal conditions in terms of methodological and substantive economy conditions. The question that arises is whether these are sufficient to construct empirically viable accounts of UG. The aim of the enterprise is to see how far one can get exploiting just these considerations.

Let's begin with a consideration of (B) above, the four levels hypothesis. GB identifies four critical levels in the structural analysis of a sentence: its DS, SS, LF and PF forms. Why four levels? From a minimalist perspective, LF and PF are unobjectionable. Recall, that one of the big facts about natural languages is that they pair sound and meaning. LF and PF are the grammatical inputs to the conceptual and sound systems respectively. As any adequate grammar must provide every sentence with a sound and a semantic interpretation, any adequate grammar must thus have an LF and PF representation. In this sense, LF and PF are conceptually necessary parts of any adequate model of grammar.

What of SS and DS? Let's consider these in turn, SS first.

(i) SS, Arguments from Case Theory and Variation:

SS is a theory internal level. This means that it is not motivated by the general sorts of considerations outlined in chapter 1. Thus, the motivation for SS is empirical, not conceptual. This, it is important to emphasize, is NOT a criticism. It is merely an observation that points to another question; namely how strong is the evidence for postulating SS? What empirical ground would we loose if we dropped the assumption that SS exists?

On the face of it, we would loose quite a bit. First, within GB, both case and binding theory apply at SS as does gamma marking, in various "Barriers" versions of GB. Second, SS serves an important descriptive function in that it marks the border between overt and covert syntax. As much language variation has been treated in terms of rules applying before or after SS, it would appear that dispensing with SS would leave us without the descriptive resources to describe this variation. Lastly, there are various kinds of phenomena that seem tied to SS. For example, parasitic gap licensing. So, it would appear that SS has considerable empirical value, even if it is conceptually unmotivated.

The minimalist project is clear: to show that appearances here are deceiving and that it is possible to cover the same (or more) empirical ground without the benefit of SS. This is what Chomsky (1993) tries to do. Let's review his reasoning.

First, what is theory of case and what reasons are there for thinking that it applies at SS? Case is assigned to a D/NP by a V or P under government. D/NPs must be assigned case to be well formed (the case filter). It is assigned at SS. Why there? Because case has been argued to be relevant at both LF and PF and not to be relevant at DS. (5) shows that case cannot be assigned at DS. Consider passive sentences like (5a). At DS, the D/NP that surfaces as nominative is in the object position-(5b). It is not governed by the finite tense given that it is complement to the passive V participle, a non-case assigner. Only after movement to Spec IP is it governed by Tns and so case markable. -(5c).

(5) a.   He was arrested
     b.   [IP 0 Tns was [VP arrested he]]
     c.   [IP He$_i$ Tns was [VP arrested t$_i$]]

The same argument can be made in raising constructions. (6a) has the DS (6b). The object is not governed by the matrix finite T until it occupies the SS position (6c).

(6) a.   He seems to be likely to win
     b.   [IP 0 Tns [VP seems [IP 0 to [VP be likely [IP he to win]]]]]
     c.   [IP He$_i$ Tns [VP seems [IP t$_i$ to [VP be likely [IP t$_i$ to win]]]]]

Note that these same cases argue that case assignment takes place AFTER movement. In other words, the case that a D/NP displays is related to the position to which it moves, not the one from which it came. Why then not assume that it takes place at LF? Why SS? The main reason is that LF operations, recall, are phonetically inert. However, what case a D/NP receives quite clearly has phonological implications. English pronouns surface as 'he', 'she' etc. If nominative but as 'him', 'her' etc. if accusative. In other words, what case one gets has phonological implications. In a T-model, this means that case cannot be assigned at LF.

What about PF? Here the argument is somewhat more theory internal. Most late versions of GB assume that case theory and theta theory are linked by the Visibility Condition.

(7) Visibility Condition:
     A D/NP's theta role is visible only if it is case marked.

(7) reduces the case filter to the theta criterion as, with (7), the theta role of a non-case marked D/NP (chain) is invisible. Hence a D/NP without case violates the requirement that all D/NPs must bear theta roles.

The main motivations for so reducing the case filter revolves around existential constructions. The details are irrelevant here.[See note] Let's assume that being case marked is a precondition for theta visibility. If so, case marking must take place after DS, feed PF and feed LF. SS is the level that meets all three requirements and so seems to be the appropriate locus for case assignment.

In short, it appears that the GB theory of case requires the use of SS. And this is very good argument for the existence of SS given the strong empirical evidence in favor of case theory. However, appearances are deceptive here. Chomsky 1993 shows that the conclusion above crucially rests on an unwarranted *technical* assumption about how case is implemented within

GB and that if we adopt slightly different (but no less adequate) technology then the need for SS disappears. In particular the above arguments rest on the assumption that case is ASSIGNED. It behooves us to consider what case assignment is.

A head V/P/finite-T assigns case in the following sense. D/NPs are inserted into D-structures without case features, i.e. they are case-nude. There is an operation where a head marks this case-nude D/NP with a case; nominative if finite-Tns, accusative if V or P. The mechanics of standard GB case theory assumes (a) that on lexical insertion D/NPs have no case and (b) that case is acquired through the course of the derivation. With this technology at hand, we have argued above that case must be discharged by SS in order to be empirically adequate.

However, why assume that this is the way that case works. What, for example, if we assumed that case was checked rather than assigned. A checking account of case involves the following technology: D/NPs have case features at DS. The appropriateness of these features is checked derivationally. Thus, an accusatively marked D/NP, for example, checks its case against a V. If there is no V around, then these features cannot be checked and the derivation is cancelled. In place of a case filter saying that D/NPs must bear case, we substitute a version that requires the case features of a D/NP to be checked. In place of assignment, we substitute checking.

There is little empirical reason for preferring case assignment to case checking. However, and this is the surprise, if we assume that case is checked rather than assigned then the above arguments in favor of SS evaporate. In later chapters we revisit case theory from a minimalist perspective and change some fundamental assumptions of the GB approach. However, the present argument does not rely on any major revisions of case theory. It only relies on substituting checking for assignment. All else is left in place. Chomsky's point is that this trivial technical emmendation suffices to undercut the case based arguments in favor of SS. Consider the details.

Recall that the main argument in favor of the claim that we needed to check case at SS rather, than say LF, was that case is phonetically visible; 'he' differs from 'him' and 'she' from 'her'. For this reason we could not assume that case was an LF process. However, with our proposed new machinery, this problem evaporates. All D/NPs already have their case features at DS. But if so, getting something on the relevant D/NPs for purposes of pronunciation is not a problem. All we need to be sure of is that the right case appears in the right place, e.g. that 'he' appears in subject positions of finite verbs and not the object position of transitive verbs. However, this sort of checking CAN be delayed until LF at no empirical cost. So, if we replace assignment with checking and assume that the case filter applies at LF (by LF all cases must be appropriately checked) then all goes swimmingly even without SS.

Consider a couple of concrete examples to see that this is indeed so.

(8) a. * Mary to leave would be terrible
     b.   John loves they
     c.   It was seen them

On an assignment story, (8a) is out by the case filter. The infinitival Inflection assigns no case and so 'Mary' is caseless violating the filter. On the checking story, 'Mary' has a case on insertion but there is nothing to check this case as the only governing head is the non-finite Inflection and this is not a case active head. As such the case cannot be checked and 'Mary' violates the case filter at LF by having an unchecked case.

The same story extends to (8c). The passive verb 'seen' is not a case assigner, nor a case checker. As such, 'them' cannot get case nor check case against it. Violation of the case filter follows.

(8b) is a little different. Here 'they' is the wrong case, nominative instead of accusative. On the assignment story, this follows because V only assigns accusative case and 'they' is governed by V. Similarly, we can assume that V in (8b) checks accusative case. However, this results in a case clash with the nominative case of 'they'. On the natural assumption that case clashes are disallowed, this "checking" results in ungrammaticality.

In sum, we cover virtually the same empirical ground on either approach. However, with checking in place of assignment, we can assume that the case filter applies at LF and we can dispense with any mention of SS. What this shows is our earlier arguments in favor of SS rested on a technical implementation that is easily avoided and that these sorts of arguments should not stand in the way of the minimalist project of doing away with SS.

There is a second kind of argument advanced in favor of SS that yields to a similar technical reworking of the theory. It is well known that languages differ in many respects in their overt properties. For example, WH-questions in English are formed by moving Wh's to the front while in Japanese these Wh's remain in place.

(9)   a.   What did Bill see
       b.   Bill what saw

Or languages like French raise verbs to finite T overtly while in English these verbs stay in place.

(10)   a.   FR John sees often Mary
       b. * John sees often Mary

The way these differences are managed in GB is to say that Japanese does covertly what English does overtly and that English does covertly what French does overtly. In other words, a standard assumption is that all languages are identical at LF and that the overtly moved cases tell us what all languages "look like" at LF. Assume that this is indeed so. This seems to demand that we be able to make a distinction between overt and covert movement. Moreover, if we assume something like the T-model, then this assumes that there is something like SS as this level defines the border between overt (phonetically altering) movement and covert (phonetically impotent) movement.

Once again, it is questionable whether this line of argument actually establishes the need for a level that distinguishes overt from covert movement. Consider the following alternative: assume that movement is driven by the need to check features, as in the case example above. And assume that features come in two flavors, weak and strong. Strong features are phonologically indigestible and so must be checked before the grammar splits. Weak features are phonologically acceptable so need only be checked by LF. Assume further, that grammars are "lazy" in that one does not check features unless one must. Then, given that weak features are phonologically fine, they will not be checked until LF. Of course, if strong features are no checked before the grammar splits the derivation will phonologically gag. So strong features will be checked by overt movement. We can now say, the differences noted among languages is simply a question of feature strength. Consider how this works with the example above.

We say that in English WH elements must check a strong feature in C while in Japanese only a weak feature needs checking. Say, for the sake of discussion, that this feature is actually a

feature of C0, a strong WH feature in comp. This is checked against a WH in Spec CP. This forces, a WH element to overtly move in English as in (11).

(11) a.   [C+Strong WH [......WH......]]
      b.   [WH [C+Strong WH [.....t.....]]

In Japanese, in contrast, this WH feature in C is weak and so is checked only after the grammar splits, I.e. Covertly. Note that this (more or less) suffices to get the difference between English and Japanese, all without mentioning SS. Thus, a technology of Strong versus Weak features is all that we need to accommodate variation. Why? Because contrary to first impressions, the overt/covert distinction which descriptions of parametric variation exploit do not require invocations of SS. HW: show how V-raising in English versus French would be handled in a feature checking approach.

A question worth asking is whether this use of features is better or worse than the earlier GB treatment in terms of rules that apply before and after SS. I do not believe so. Ask why it is that English Wh's are moved overtly while Japanese ones are moved covertly. There is no principled account. That's just the way that things are. So, within the more standard GB account we had no account for WHY some operation occurred prior to SS in one language and post SS in another. Given that this is so, invoking features leaves us no worse off than assuming that some operations are pre-SS and some post-SS. Does it leave us better off? Yes and no. There is nothing particularly principled about an account based on features. They are too easy to postulate and thus carry rather little explanatory power. However, in the present context the feature based approach tells us something interesting: that variation provides no evidence for a level like SS. The reason is that we can deploy technology that is no less adequate and no less principled but that does not need SS at all. This is an interesting conclusion for it suggests that SS is an artifact of our technical implementation rather than a level supported on either strong conceptual or empirical grounds. Given this, we are free to consider discarding SS.

One last point. Doesn't the overt/covert distinction require SS at least in the following sense: the grammar splits and SS is the place where this happens. Without splitting the distinction between movements that affect phonology and those that do not could not be sustained. SS is needed as the point in a derivation where the split occurs.

The argument above presupposes that the T-model requires a level like SS. However, though SS suffices to give one the T-model it is not necessary. To make discussion concrete, let's assume (with Chomsky 1993) that at some point in the derivation rule of SPELL OUT applies that rips the phrase marker into, sending phono-relevant info to PF and semantically relevant info to LF. SPELL OUT is another way of implementing the idea of a T-model. Now, postulating SS amounts to saying that there is a point in every derivation where SPELL OUT applies, namely SS. However, the T-model is consistent with a weaker claim: that in every derivation SPELL OUT applies at some point, not necessarily the SAME point in every derivation. This weaker version suffices to give one a T-model, but makes no claims about a unique spot in all derivations where splitting applies. This weaker version allows an overt/covert distinction without committing hostages to an SS level.

What's the upshot here? We have reviewed GB arguments in favor of the idea that SS is required. These empirical arguments, we have seen, only go through on the basis of certain technical assumptions that are of dubious standing. If we replace these with other implementations, we are left with accounts no less empirically adequate than the standard GB accounts but without an SS level. This suggests that the standing of SS in GB is less empirically solid than generally believed. Of course, more needs to be addressed before this conclusion is

embraced: there are other phenomena having to do with parasitic gap constructions, binding phenomena, and subjacency effects. We return to these below. What we have hopefully shown, however, is that it is not OBVIOUSLY empirically hopeless to try to eliminate SS.

(ii) Arguments from Binding Theory:
There are another set of arguments for SS from the binding theory that Chomsky (1993) reviews. Let's outline these here.
Consider (12).

(12) a. Which picture that Harry bought did he like t
       b. He liked this picture that Harry bought
       c. Which man said he liked which picture that Harry bought

The relevant data involves the possible coreference between 'he' and 'Harry' in these sentences. It is possible in (12a), impossible in (12b,c). The data in (12b,c) can be attributed to principle C if we assume that it applies at SS. Recall, that Principle C forbids an R-expression from being A-bound. If we assume this, (12b) is fine because 'Harry' is not c-commanded (and so not bound) by 'he' at SS. (12B,c) are out because at SS 'he' does c-command 'Harry' and so co-indexing is forbidden. The fact that (12c) patterns like (12b) argues that SS is the relevant level if we assume, as most GB analyses did, that at LF the structure of multiple interrogatives involves movement of the in-situ WH phrase to the phrase containing the WH in Spec CP. Thus, at LF, the structure of (12c) is analogous to that of (12a) with the WH phrase headed by 'which picture' in the matrix Spec CP.

(13) [Which man, which picture that Harry bought [t1 said he liked t2]]

Observe that (13) is structurally similar to (12a) as regards the c-command relations between 'he' and 'Harry'.
In sum, if we assume that principle C is checked at SS, then it appears that we have an argument for SS in terms of the Binding Theory.
However, once again, appearances are somewhat deceiving. Note that the above argument relies on the assumption that at LF the structure of (12c) is (13), i.e. that at LF WH-raising moves the whole WH phrase to CP. If, in contrast, we assumed that ONLY the wh element moves at LF, to establish a structure sufficient for multiple interrogative interpretation, then at LF the structure of (12b) is (14), a structure in which principle C would force 'he' and 'Harry' to be disjoint.

(14) [Which man1,which2 [t1 said he liked t2 picture that Harry bought]]

Thus, the argument above for SS is as good as the supposition that at LF the WHOLE WH phrase is moved to the CP occupied by the other WH. What then are the arguments for this?
As it turns out, the arguments are quite weak. Recall, the main purpose of moving the in-situ WH to the WH filled CP is to establish a relation between the Wh's so that the paired reading results. Even if one assumes that this requires LF movement, it is not clear that it requires moving the WHOLE WH expression rather than just part. As such, there is little reason to think that the premise underlying the argument is correct.
Aside from the observation that in overt syntax one can move the whole WH phrase, there is little reason to think that at LF one does so. In fact, in overt syntax, it is not clearly necessary to move the whole WH.

(15)   Which portrait did she buy that Mary likes

In (15), the relative clause has not moved with the rest of the WH and we find that 'she' and 'Mary' cannot be interpreted as co-referential. Note that (14) and (15) are structurally quite similar at LF and this would account for the inability of co-reference in (13c). We need a little more, however, we need to prevent "pied piping" the relative clause at LF. We can enforce this in various ways: e.g. Assuming that LF moves as little as it can get away with, or it is preferable at LF to modify variables rather than operators (all things being equal). At any rate, it seems possible to defuse the premise that is causing the problems without too much trouble so it is safe to conclude that this binding theoretic argument for SS is weak at best.

There are other considerations that favor postulating an SS level. We return to a discussion of some of these below. In particular there are some constructions that seem to be licensed at SS (e.g. Parasitic gaps). We return below to show that these arguments too can be set aside but more technical apparatus will be needed before we tackle this issue.

So, a conclusion: we have seen that there are methodological reasons to hope that SS does not exist: it is not a conceptually required level. Moreover, we have seen that the empirical evidence for SS is weak, at best and that some technically feasible alternative implementations of the leading ideas of GB results in a version of UG where SS can be dispensed with. This provides hope that the methodologically best theory is also not too far removed from empirical adequacy. This is what we have seen.

Second point. The argument till this point has been very conservative. We have taken the conceptual architecture behind the GB apparatus largely at face value and seen that small technical changes allowed us to remove what appeared to be a deeply entrenched architectural property, namely the postulation of an SS level. Later on we will suggest more radical revisions of GB. However, it is surprising how salutory thinking the details afresh has been just for our appreciation of GB itself.

(iii) DS: the arguments for this level.
Let's begin as we did when considering SS: how is DS characterized within GB? It is the level where lexical properties meet the grammar, the grammatical realization of pure GF-theta. Logical objects are syntactic objects at this level, logical subjects, syntactic subjects etc. The two core modules that work at this level are the theta module and the phrase structure module (X'-theory). DS is the locus of grammatical recursion precisely because of this. Putting adjunction to one side, sentences can be of arbitrary length precisely because a category A can be embedded within another category of type A. Thus, for example, we find sentences embedded within sentences and D/NPs within D/NPs. Once this sort of embedding obtains, there is no upper bound on sentence length. Adjunction is another process that yields the same unboundedness. Within GB, adjunctions are also represented at DS. In fact, sentence size, within GB, is largely a function of the DS a sentence has. Movements do not (generally) enlarge sentences, nor do construal processes.

Functionally, DS is the output of phrase structure operations and lexical insertion. It is the input to overt movement operations. Substantively, it is the phrase marker level characterized as follows: all and only thematic positions are filled.

In sum, then, we can characterize DS as

(i) the level that grammatically realizes lexical properties (all and only) in X' structures.
(ii) output of phrase structure and lexical insertion operations
(iii) the input to overt movement operations

When we ask if DS exists, or if it is required, we are asking whether there is a need for a level of grammar meeting these three requirements.

In order to focus on this question we need to determine the empirical utility of postulating a level like DS: what does having DS buy us empirically? In our view, the main empirical consequence of adopting DS is that it accounts for the difference between raising and control structures. So, let's consider the properties of these two types of constructions and see how it is that DS "derives" them.

(a) Control versus Raising
Consider an example of each

(15) a.   John seemed to leave early
      b.   John hoped to leave early

The first difference between the two examples is semantically palpable. In (15b) John plays two roles. He is a hoper and a leaver. In (15a), John is a leaver but there is no sense in which he is a seemer. This is seen in the fact that (15a) has an adequate paraphrase as (16).

(16) It seemed that John left early

No analogous paraphrase exists for (15b). (17) does not convey the same thing as (15b).

(17) It was hoped that John would leave early

Technically, these facts are described by saying that 'John' has two thematic functions in (15b) but only one in (15a). John is the logical subject of both the matrix and the embedded verb. In (15a) it is the logical subject of the embedded verb alone. In a word, there is no theta role that John has in virtue of being the syntactic subject of the matrix clause.

These thematic differences are further evidenced by the fact that the subjects of raising verbs can be filled by non-theta bearing expressions. For example, epithets are acceptable with 'seem' but not 'hope.'

(18) a.   It seemed that John left early
      b. *   It hopes that John leaves early
      c.   There seemed to be a man at the party
      d. *   There hoped to be a man at the party

Also, we can find idiom chunks in the non-thematic subject positions of raising but not control predicates.

(19) a.   The shit seemed to hit the fan
      b.   All hell seemed to break loose
      c. *   The shit hoped to hit the fan
      d.   All hell hoped to break loose

On the assumption that idioms cannot bear regular theta roles, theya re barred from theta positions. This accounts for (19c,d). Given that the subjects of raising predicates are non-thematic, it also permits examples like (19a,b).

A third difference between raising an control arises when we embed active/passive pairs within the two kinds of predicates.

(20) a. The Dr seemed to examine John
      b. John seemed to be examined by the Dr
      c. The Dr hoped to examine John
      d. John hoped to be examined by the Dr

Whereas (20a,b) are tolerably good paraphrases of one another (both are true in the same contexts), (20c,d) are clearly very different in meaning. The reason is now clear. The only thematic role 'the Dr' or 'John' has in (20a,b) is the one that it gets in the embedded clause. As active/passive pairs have the same thematic roles (e.g. In "John saw Mary" and "Mary was seen by John" 'John' and 'Mary' play the same roles in both sentences) and the matrix predicate adds no thematic information, (20a,b) are thematically identical and this accounts for their essential synonomy. In (20c,d), in contrast, the matrix verb 'hope' has a thematic subject. Thus, in (20c), 'The Dr' has the hoper role (in addition to the examiner role) while in (20d) this role is assumed by 'John' (in addition to the examinee role). This makes for semantically very different propositions and hence the felt difference in meaning under passivization.

In sum, these kinds of data support a thematic distinction between raising and control verbs. Within GB this difference is represented in terms of a null anaphoric element PRO. PROs are empty categories that can be base generated in thematic positions. They are bound (or "controlled") by higher nominal expressions. For example, (15b) above has the DS (21).

(21) John hoped [PRO to leave early]

Note that PRO occupies the external argument position of 'leave' and 'John' the external argument position of 'hope'. By LF, 'John' and 'PRO' become co-indexed and this accounts for the observation that John appears to play two roles in a sentence like (15b).

Raising structures contrast with control configurations as follows. At DS a sentence like (15a) has the structure (22a). 'John' raises to the matrix subject position (for case reasons) yielding the SS (22b). The position to which 'John' raises is non-thematic. The movement leaves behind a co-indexed trace.

(22) a. [seemed [John to leave early]]
      b. [John1 seemed [t1 to leave early]]

Note that traces (e.g. 'T1') are the products of movement. PROs, in contrast, are expressions that are inserted at DS.

So, to recap, within GB, raising and control predicates are structurally distinct. Control predicates involve PROs that are bound by antecedents in thematic positions. Raising predicates involve (NP-)traces that are bound by antecedents in non-thematic positions. Observe that the differences observed above are accounted for with these two different configurations. For example, the reason that expletives are barred from the subject positions of control predicates is that they are thematic positions. They are permitted in the subjects of non-thematic predicates because these are not thematic. This same difference accounts for the voice transparency contrasts noted above.

So, given these structural differences we can account for the differing properties of control and raising predicates. The virtue of DS is that its properties allow us to derive these structural differences from the lexical differences of the two kinds of predicates. Let's see how.

Recall that DS is the pure representation of thematic properties in phrasal garb. This means that all lexical/thematic properties are satisfied there and only thematic properties are satisfied there. Consider what this means for a control predicate like 'hope' in (21). 'Hope' is a verb that takes a propositional complement (the state hoped for). It also has an external argument (a 'hoper'). Given the notion that DS must have all thematic properties discharged, this means that at DS, a sentence involving 'hope' must have its subject position "saturated" and also have a sentential complement. The embedded verb of the sentential complement will also have a theta "grid." Say it is the two place verb 'kiss'. It requires a 'kissee' and a 'kisser' argument at DS. So the structure of (23) must have the positions indicated by ß filled at DS as these are thematic positions and all such positions must be filled for a DS to be licit.

(23) [ß hopes [ß kiss ß]]

Consider in this light (24).

(24) Mary hopes to kiss John

Here, 'Mary' is the hoper and 'John' the kissee. Note that we require an kisser for the embedded clause. Clearly this argument is not phonetically realized. We know from the meaning of (24) that this element is related to 'Mary' in some way. However, by the definition of DS it cannot be a trace as DS is prior to movement. The only ec that it can be, the only ec that can be base generated is a PRO, and by the definition of DS a PRO must sit in the external argument position of the embedded clause. Its DS MUST be (25).

(25) Mary hopes [PRO to kiss John]

Observe, that the GB view of DS and the necessity for an expression like PRO are very intimately connected. In effect, the requirements of DS force the postulation of null categories like PRO, ec's not formed by movement, given the plain fact that verbs can take non-finite complements.

Consider now what DS imposes on verbs like 'seem' when they take non-finite complements. The subject position of 'seem' is non-thematic. This means that at DS NOTHING can be in the DS subject position of this predicate (or at least nothing thematic). If the embedded complement clause is transitive, then the embedded verb must take a pair of arguments. For example, if the matrix is 'seem' and the embedded clause is 'kiss' only the positions indicated by ß can be filled at DS.

(26) [seem [ß kiss ß]]

A sentence like (27) must then have the DS (28).

(27) Mary seemed to kiss John

(28) [seem [Mary to kiss John]]

13

Note that to derive (27) from (28) movement is required (case reasons force 'Mary' to move to the matrix). This yields the SS structure (29) for (27).

(29)   Mary$_1$ seemed [t$_1$ to kiss John]

Note that ONLY a trace can surface in this position given the DS (28). More to the point, PRO is excluded from this position in (29) just as trace is from the embedded subject position in (25).

Recall that the contrast between raising and control reduces (in GB) to whether the ec in the embedded subject position, the ec related to the syntactic matrix subject, is a PRO or an NP-t. This is precisely the difference that DS motivated in raising versus control verbs. We now have all that we need to derive the data noted earlier. Consider this.

Expletives are forbidden from the subject positions of control predicates because these are theta positions (see (23)). As expletives do not have the semantic heft to bear theta roles, they cannot be generated in theta positions.

The idiom material follows as well if we assume that idioms are lexically determined, viz. Lexical facts and so determined at DS. A sentential idiom like "the shit hit the fan" must form a unit like (30a) at DS. Note that this is compatible with raising to the matrix subject position as this is not filled at DS, (30b). It is prevented in control structures as "the shit" is not locally related to "hit the fan" if generated on high and so cannot have an idiomatic reading, (30c). Note too, that on the idiomatic reading there is little reason to believe that "the shit" can bear a theta role.

(30)   a.   [The shit hit the fan]
       b.   [seems [the shit hit the fan]]
       c.   [the shit hoped [PRO to hit the fan]]

Last comes voice transparency. We explain the voice transparency of (31a,b) by noting that movement does not affect thematic structure.

(31)   a.   The Dr seemed to examine Mary
       b.   Mary seemed to be examined by the Dr

In particular, the active and passive embedded clauses have the same thematic structure. This is seen in the near paraphrases in (32).

(32)   a.   It seemed that the Dr examined Mary
       b.   It seemed that Mary was examined by the Dr

Furthermore, raising the embedded subjects in (31a,b) has no thematic effect. Hence there is voice transparency.

Now consider (33a,b). Their corresponding DSs are very different, (34a,b).

(33)   a.   The Dr hoped to examine Mary
       b.   Mary hoped to be examined by the Dr.

(34)   a.   [The Dr hoped [PRO to examine Mary]]
       b.   [Mary hoped [PRO to be examined by the Dr]]

Observe that in (34a) 'the Dr' has the 'hoper' role while in (34b) 'Mary' does. This is sufficient to explain why (33) lacks voice transparency.

In sum, given DS, we derive the properties characteristic of raising versus control structures. And this IS a big deal. The question we turn to next is whether we need DS to do this or if there is another way.

(b) Doing Without DS

The distinction between raising and control that DS captures rests on the theta theoretic distinction between raising and control predicates. DS mediates these thematic facts by mapping principles that are deemed to hold at DS. The question that we need to address now is whether DS is required to state these thematic requirements. In other words, is a DS level needed or is it just a useful device? Chomsky (1993) suggests that it is not required and that, in fact, DS as conceived by GB is quite problematic. We first review what we can put in the place of DS.

Assume that we dump DS. What do we need to do? One thing we need to do is find a place for the generative engine of the grammar. Recall that within GB the locus of generativity is DS. We get this property because expressions of type A can embed within expressions of type A (noun phrases within noun phrases or sentences within sentences). This sort of embedding allows recursion and so no upper bound on sentence size. As noted above, the locus of this recursion is DS.

However, DS is not required for recursion. In fact, in earlier approaches to UG did without DS entirely. In its place were rules that combined lexical atoms to get bigger and bigger structures. We can revert to this sort of theory and thereby arrive at a theory that seems to do without DS. Let's see how.

Say that we have a lexicon where lexical atoms are housed and grammatical rules of various types that put these words together. Assume that one of these rules of combination corresponds to phrasal rules that require that when words combine they respect X' structure. For example, consider (35).

(35)   Bill saw Mary

To construct this sentence we take 'see' and 'Mary' from the lexicon. We combine them by projecting 'saw' and MERGING it with 'Mary'. To receive a theta role, 'Mary' must be in the immediate projection of 'saw'. This give us something like (36).

(36)   [VP see Mary]

We proceed by taking PAST out of the lexicon and combining it with the VP to get (37).

(37)   [T Past [VP see Mary]]

We then pull 'Bill' out and combine it to get (38).

(38)   [TP Bill [T PAST [VP see Mary]]]

To this point, we have simply constructed the sentence bit by bit bottom up. In what sense is this not simply a coding of DS? There are two ways in which this might be different. First, so building a sentence need not imply that phrase structure processes and movement operations are segregated. Recall, that in GB ALL PS rules precede ALL movement operations. In the

15

constructive account above this need not be true. In fact, we will assume that it isn't. Consider the following derivation as an illustration.

(39) I wonder what Bill ate

First one puts together 'eat' and 'what' to give you '[VP eat what]'. This is then combined with 'Past' to get '[T Past [VP eat what]]'. Then we merge 'Bill' to get the clause (=TP) '[TP Bill [T Past [VP eat what]]]'. Next we combine a +WH C to get '[C C0{+WH} [TP Bill [T Past [VP eat what]]]]'. Then 'what' is moved to Spec CP to get [CP What [C' C [TP Bill [T Past [VP eat t]]]]'. Then this is combined with 'wonder' and the result of this is combined with Present tense and 'I' to give you (40).

(40) [TP I [T Pres [VP wonder [CP What [C' C [TP Bill [T Past [VP eat t]]]]]]]]

The important thing to note about this derivation is that it combines Merge ("put together") phrasal operations with Movement operations in the course of a single derivation. This is what a GB approach that implements a DS level forbids. In effect, DS enforces a condition that says one must satisfy ALL lexical properties within a phrase marker BEFORE any movements can take place. The derivation of (40) above relaxes this requirement. By so doing, the derivation rejects the strictures of DS.

So, it is possible to have generativity without DS. What about the interesting property; namely the theta requirements that give us the distinction between raising and control? Is DS needed for this. Not really. We can get the distinction we want without a DS level by making it impossible to get a theta role by movement. In other words, say we adopted a principle like (41).

(41) Theta roles can only be assigned by the MERGE ("put together")   operation

If we assume that this is so, then the raising versus control facts that we observed above follow as well. Note that (41) is not stated on any level of representation. Rather, it is a condition on grammatical operations. It permits assignment of thematic relations by the MERGE operation that puts phrasal units together and prohibits thematic assignment via the MOVE operations that displace phrases. So reading (41) should sound familiar. It echoes the DS requirement that thematic requirements must be exhausted at DS viz. That all and only thematic positions are filled at DS. The DS requirement had as an effect that movement could not satisfy thematic requirements and that only lexical insertion prior to movement could. (41) says exactly the same thing, all without adverting to a level like DS, by placing the required thematic restrictions on the MERGE and MOVE operations instead of a DS level. We return later on to discuss how (41) is executed within various minimalist approaches and consider the virtues of adopting it. For current purposes it is sufficient to see that (41) will do whatever DS did all without assuming that we need a level like DS that acts to strictly segregate movement operations from phrase structure rules.

For the sake of illustration, consider once again voice transparency. (42) is derived by as follows.

(42) The Dr seemed to examine Mary

We first combine 'examine' and 'Mary' to get 'examine Mary'. We then combine 'to' with the VP to get 'to [VP examine Mary]'. Then the embedded subject is merged. Note that the merging of 'Mary' and 'The Dr' in these positions coincides with an assignment of a theta role as this is merging into theta positions. The result is '[IP The Dr to [VP examine Mary]]'. We

then add 'seem' and 'PAST' to get '[T Past [VP seem [IP the Dr to [VP examine Mary]]]]'. At this point, we move 'the DR' to matrix position (to get case) and we end up with '[TP the Dr [T PAST [VP seem [t to [VP examine Mary]]]]]'.

Note that the structure derived here looks exactly (32c). (32d) would be analogously derived but with a movement occurring in the embedded clause for passive.

Now consider a non transparent example, (43).

(43)   The Dr hopes to examine Mary

Here, the derivation proceeds as follows. Again we start by combining 'examine' and 'Mary'. This merger gives 'Mary' the examinee theta role. Then 'to' is added to give 'to [VP examine Mary]'. Then what? We need to merge something here to discharge the external examiner theta role. If we merge 'the Dr' we run into trouble as then we need to move 'the Dr' to the higher clause. But recall that (41) only permits theta discharge via Merge. It forbids such discharge via Move. So, IF we merge 'the Dr' here we will be stuck. What can we do? We merge PRO to get '[IP PRO to [VP examine Mary]]'. On so merging PRO, PRO is receives the examiner theta role. Then we add 'hope' and 'PRES' to get '[T PRES [VP hope [PRO to [VP examine Mary]]]]'. Then we merge 'the Dr'. This merger gives 'the Dr' the hoper role. At some point in the derivation, 'the Dr' and 'PRO' are co-indexed. This results in reading in which we understand the DR to be both a hoper and examiner. The resulting structure is '[TP The Dr [T PRES [VP hope [IP PRO to [VP examine Mary]]]]]'. Note that this is identical to (34a) above. The derivation of the passive sentence (44) would proceed analogously but this time 'Mary' would be merged in the matrix as subject of 'hope' and this would result in Mary having the hoper and examinee role. This explains the lack of transparency in these cases.

The other properties discussed are similarly explained. The reason is not hard to spot. (41) functions in a derivational system exactly like the thematic requirements that DS imposes function in GB. It turns out that the DS level is not actually required to handle the contrast between raising and control structures. It is sufficient but not necessary. What is required to mimic the effects of GB is to restrict theta discharge exclusively to merge operations. Once this is done (and (41) does precisely this) it adds nothing to order all merge operations before any movement operations as DS does.

In fact, Chomsky argues that DS is actually inadequate empirically. The problem arises with constructions that discharge theta roles in slightly unorthodox ways. A particularly problematic construction for DS as conceived by GB comes from a consideration of 'tough'-constructions like (44).

(44)   Moby Dick is hard for Bill to read

Let's consider some properties of this construction.

First, 'this book' is in the subject position of the matrix clauses (check out the agreement between 'this book' and 'is'). Second, the theta role that it has is that it is the thing read. In other words, it is related to the object position of 'read'. Third, the matrix subject position appears to be a non theta position given the paraphrase (45).

(45)   It is hard for Bill to read Moby Dick

Thus, fourth, 'Moby Dick' must be related to the object position via movement. The problem is that it is quite unclear what sort of movement this could be.

The standard ways in which a D/NP in a non-theta position receives a theta role is by A-moving from the position in which its theta role is assigned. In (44) A-movement, however, is impossible. The structure would be (46).

(46) [Moby Dick is hard [for Bill to read t]]

Recall that NP-t is subject to a locality condition like Principle A. The domain of the trace in (46) is the embedded clause. It is unbound there and so the structure should be filtered out. Another problem is case theoretic. A movement is generally driven by case requirements. In (46), 't' is in a case position. This leaves the motivation for movement somewhat up in the air (especially once one goes in a minimalist direction). At any rate, it seems that whatever sort of movement we have here it cannot be A-movement.

Chomsky (1981) suggested that this is actually an instance of A' movement with a 0-operator moving to the around 'hard' and forming a complex predicate with 'hard'. The structure looks something like (47) with 'hard [0 [for Bill to read t]]' a kind of complex predicate predicated of the subject 'Moby Dick'. Under predication, 'Moby Dick' receives a theta role. In effect, then, the subject position in (44) IS a theta position.

(47) Moby Dick is [AP hard [0 [for Bill to read t]]

Now complex predicates are not quite as exotic as they may appear to be. We find them in relative clauses where a sentence can function as a kind of giant adjective. Consider (48).

(48) John read a book that Bill enjoyed

Here, 'a book that Bill enjoyed is a constituent'. It has the readee theta role. Moreover, 'a book' also has the enjoyee theta role. We know that relative clauses are formed via WH/A'-movement. So it is possible to get what looks like theta assignment with the help of an A'-movement operation. The proposal in Chomsky (1981) is that the same thing happens in 'tough' constructions with the difference that the adjective and its complement form a complex predicate. Let's assume that this account is on the right track and ask what this implies for DS.

Well, it is clearly problematic. Observe that the subject position is only a theta position AFTER A'-movement has taken place (recall (45) shows that the subject is not ALWAYS a theta position). However, when then is 'Moby Dick' inserted? If at DS, then it is not inserted AT THE POINT THAT THE SUBJECT IS A THETA POSITION. If after the 0-operator has moved then we can have insertion into a theta position AFTER DS. Either way, these constructions create a tension between the two leading claims of DS; that it precedes all movement and that all and only theta positions are filled at DS.

Chomsky (1981) "solved" this problem by weakening the theta requirements on DS to allow a lexical item like 'Moby Dick' to be inserted in the course of the derivation and get its theta role assigned at LF. In effect, lexical insertion and theta assignment are pulled apart, at least for atomic lexical items.

The problem is that it is not only atomic D/NPs that can appear in these positions. The subject in (44) can be of arbitrary complexity.

(49) MD being hard to read is tough for Bill to understand

Here the subject is hardly atomic. Moreover, the subject itself requires 'tough' movement to operate in its very formation. So, there is no upper bound on the complexity of what can be treated as the subject in these constructions and so the "solution" fails.

Interestingly, there is no problem if we dispense with DS and adopt (41). Recall, that we can intersperse Move and Merge operations here. Thus, we can have 0-operator movement apply before merging the subject up top. This allows us to derive these constructions with no difficulty. Consider a derivation of (47) for illustration. We first merge the 0 operator and 'read'. We continue to merge successively until we derive (50).

(50)   [for Bill to read 0]

We then Move '0' to CP to get (51).

(51)   [0 [for Bill to read t]]

(51) then merges with 'hard' forming the complex predicate, a predicate that assigns a theta role to its external subject argument. We then merge 'Moby Dick'. Note, that it is being merged into a theta position. Thus it conforms with (41) and gets its theta role in the same way that all other arguments do viz. By merging with a predicate that assigns a theta role. What makes 'tough'-constructions different then is not where they discharge their thematic responsibilities but that they involve complex rather than simple predicates.

This argument from examples like (44) constitute an argument that DS was too strong. The complete separation of merge from move leads to empirical problems and so should be rejected. Thus, we now have both conceptual (minimalist) reasons for dispensing with DS and empirical arguments for the same conclusion. Moreover, we have developed technology sufficient for retaining what we took to be the beneficial features of DS by adopting (41). In effect, we can have the best sort of theory and appear to loose nothing. Not bad!

3. Some Sample Derivations (to get the hang of the system).

We have replaced DS with two kinds of grammatical operations: Merge, an operation that takes two items and combines them (conforming to X' principles) and a rule that moves an element from one part of the phrase marker to another part. We have abandoned the idea that all Merges must precede all Moves. They can intersperse. Consider the derivation of (52) to get a hang of the system.

(52)   The woman saw George

First we merge 'saw' and 'George' to get '[VP saw George]'. "George' gets its theta role via this merger. What next? Next we combine 'the' and 'woman' to get '[D/NP the woman]'. Then we combine the two subparts together to get (53).

(53)   [IP [D/NP the woman] I [VP saw George]]

Note that we did not combine 'woman' with '[VP saw George]' and then add 'the' to 'woman'. Why not? Because we are assuming that (overt) grammar operations must "extend" phrase markers when they apply. In Short, the result of operation must leave the combined parts

with the same structure that they had BEFORE the operation took place. Observe that when we combined '[VP saw George]' and '[D/NP the woman]' to get (53), (53) has as subparts/constituents both of the inputs to the merge operation. This would not have been true on the alternative derivation. Let's see.

On the alternative derivation we first derive '[VP saw George]'. Then we merge 'woman' to get (54).

(54)   [woman [I [VP saw George]]]

If we then add 'the' to (54) then we must add it to 'woman'. [Recall that we know what the correct phrase structure is and in this case 'the' and 'woman' must form a constituent]. The problem with so adding it is that this merger will "thicken" but not "extend" the phrase marker. If we assume that this is disallowed, then this derivation is forbidden. The consequence is that in building a sentence we may have several "treelets" around prior to their combining into a single big tree.

Why buy into a condition like extension? Among other things it prevents some unwanted derivations. Consider one. We want to block the derivation of a sentence like (55).

(55) *Who did you wonder why/whether John gave a book to

This derivation is blocked within GB by the assumption that the CP is filled by 'why/whether' and that this prevent further movement of 'who' through this position. However, if we allow violation of extension, then we can derive this sentence by first moving 'who' through the embedded CP and then after it has vacated this position, inserting 'why/whether' into this slot. Note that this move(namely inserting 'why/whether' after 'who' has moved to the top) is blocked by the assumption that extension conditions (overt) grammatical operations. We will return later on to other instances where the extension condition plays a useful role.

4. Conclusion

DS and SS are central features of a GB model of UG. From a minimalist point of view where we try to make do with the conceptually required levels DS and SS contrast with PF and LF in being methodologically dispensable. This chapter has reviewed the kinds of empirical evidence put forward to support SS and DS. We have seen that with some technical changes we are able to defuse these arguments and "save" the relevant data without assuming that DS or SS actually exist. We have not exhaustively reviewed all the empirical data that has been used to motivate SS or DS. However, we have taken a look at a fair sampling. It seems fair to conclude that it is reasonable to hope that eliminating DS and SS will not come at too great an empirical cost (if any). Thus, at least with respect to these issues, the minimalist goal of making do with the "obvious" (as outlined in chapter 1) is a viable project. In what follows we will assume that further problems can be overcome and investigate what other changes to GB a serious commitment to minimalist goals would entail.