

1

The Minimalist Program

Introduction

It is my opinion that the implications of the Minimalist Program (MP) are more radical than generally supposed. I do not believe that the main thrust of MP is technical; whether to move features or categories for example. MP suggests that UG has a very different look from the standard picture offered by GB-based theories. This book tries to make good on this claim by outlining an approach to grammar based on *one* version of MP. I stress at the outset the qualifier “version.” Minimalism is not a theory but a program animated by certain kinds of methodological and substantive regulative ideals. These ideals are reflected in more concrete principles which are in turn used in minimalist models to analyze specific empirical phenomena. What follows is but one way of articulating the MP credo. I hope to convince you that this version spawns grammatical accounts that have a theoretically interesting structure and a fair degree of empirical support.

The task, however, is doubly difficult. First, it is unclear what the content of these precepts is. Second, there is a non-negligible distance between the content of such precepts and its formal realization in specific grammatical principles and analyses. The immediate task is to approach the first hurdle and report what I take the precepts and principles of MP to be.¹

1 Principles-Parameters and Minimalism

MP is many things to many researchers. To my mind it grows out of the perceived success of the principles and parameters (P&P) approach to grammatical competence. Here’s the story.

The central problem for grammatical theory is how it is that kids are able to acquire grammatical competence despite the impoverished nature of the data that is input to this process. No sane person doubts that the attainment of grammatical competence is influenced by the nature of the primary linguistic data (PLD); children raised in Paris learn French and those raised in Brooklyn speak English.² However, it is also clear that the knowledge attained vastly exceeds the information available in the PLD.³ This, in essence, is what Chomsky (1986b) dubbed “Plato’s Problem,” the problem of the poverty of the stimulus. The greatest virtue of P&P accounts is that they provide a way of addressing Plato’s problem in the domain of language.

2 *The Minimalist Program*

The idea is simple. Kids come biologically equipped with a set of principles of grammar construction, i.e. Universal Grammar (UG). The principles of UG have open parameters. Specific grammars arise once these parameter values are specified. Parameter values are determined on the basis of the PLD. A language specific grammar, on this view, is simply a vector specifying the values that the principles of Universal Grammar leave open.⁴ This picture of the acquisition process is sensitive to the details of the environmental input (as well as the level of development of the child's other cognitive capacities) as it is the PLD that provides the parameter values. However, the shape of the knowledge attained is not restricted to whatever information can be gleaned from the PLD since the latter exercises its influence against the background of rich principles that UG makes available.

In retrospect, syntactic research since the mid-1970s can be seen as largely aimed at elaborating this sort of picture and demonstrating its viability. Government-Binding theory (GB) is the best known version of a P&P theory of UG. It has several distinctive features.

First, GB is modular. The grammar is divided into various subcomponents sensitive to different kinds of well-formedness requirements. There are modules for case, binding, phrase structure, movement, control, theta-structure, and trace identification. These modules are tuned to different kinds of grammatical information (e.g. case versus antecedence), exploit different kinds of rules (e.g. construal versus movement) and locality principles (e.g. binding domains versus government configurations). GB modules, in short, are structurally and informationally distinct.

One of modularity's primary virtues is that it radically simplifies the kinds of rules that grammars exploit. In place of construction specific rules (such as Passive, Raising, WH movement, and Relativization), the grammar is pictured as having very general highly articulated modules whose interactions yield the properties observed in specific constructions. The modules factor out features common to different structures and allow principle based grammars to replace rule based ones.⁵ Thus, in place of grammars with rather complex rules (i.e. rules stated in terms of complicated structural descriptions and structural changes), GB contains very simple rules whose overgeneration is curtailed by the combined filtering effects of the general principles constitutive of the various modules.

Second, GB contains a very general transformational component. It contains movement rules and construal rules which index nominal expressions to one another. As a by-product of its modular design, GB has been able to adopt a very simple movement rule: 'Move alpha'. 'Move alpha' allows any category to move anywhere at any time. The modules function to circumscribe the massive overgeneration that this very general rule inevitably leads to.

'Move alpha' incorporates a version of the trace theory of movement, viz. movement always leaves a trace – a lexically empty XP of the same category – in the position from which movement originates. For example, every application of NP-movement leaves an $[_{NP} e]$ in the launching site. Traces must be licensed. The module concerned with licensing traces is the ECP. Trace theory in concert with the ECP severely constrains the movement which a 'Move

alpha' based theory permits. A central feature of GB theories is the exploitation of traces both for purposes of interpretation and to constrain the overgeneration of 'Move alpha'.

Third, a GB grammar has four critical levels at which various conditions are applied to filter out illicit phrase markers. The levels are D-structure (DS), S-structure (SS), LF and PF. The latter two are "interface" levels and constitute the grammatical contributions to semantic and phonetic interpretation respectively. DS and SS are "internal" levels and only interact with other parts of the language faculty.

DS has several distinctive properties: (i) it interfaces with the lexicon (ii) it is the level where the thematic information specific morphemes carry is integrated into the grammatical structures that transformations subsequently manipulate (iii) it is the locus of recursion in the grammar (iv) it is input to the transformational component and (v) it is the output of the phrase structure component.

SS is the point in a derivation at which the grammatical information required by the phonology splits off from the information required for meaning. Thus, LF and PF are blind to each other and only relate in virtue of being derived from a common SS phrase marker. SS is also the locus of a variety of filters from the Binding, ECP, Subjacency and Case modules. Observe that DS and (especially) SS are the most abstract levels in UG. They are the most remote from "experience" in the sense that they are furthest removed from a sentence's observable properties, its sound and meaning.

Fourth, the central grammatical relation in GB is government. This relation is ubiquitous and appears in every module of the grammar. Government lends conceptual unity to otherwise rather diverse components. Thus, though the modules themselves may be structurally very different, using different notions of locality and different kinds of rules, still they share a degree of unity in that they all exploit the same basic relation. Theta-roles and structural cases are assigned under government, binding domains are defined in terms of government, the ECP licenses traces that are in certain government configurations with their antecedents or heads, the subjacency condition on movement is defined in terms of barriers, which are in turn defined in terms of government. In short, though the modules "worry" about different kinds of information, and use different rules and locality domains they are nonetheless organized in terms of the same basic structural primitive.

GB has been very successful in illuminating the structure of grammatical competence. Given the emphasis on Plato's problem, research has focused on finding constraints of the right sort. By "right sort" I mean constraints tight enough to allow grammars to be acquired on the basis of PLD yet flexible enough to allow for the observed variation across languages. In short, finding a suitable answer to Plato's problem has been the primary research engine and GB proposals have largely been evaluated in terms of whether they satisfactorily meet its demands. This does not mean to say that other methodological standards have been irrelevant. Simplicity, and naturalness have also played a role. However, in practice, these yardsticks of theory evaluation have been quite weak and have been swamped by the requirements of outlining principles with a reasonable hope of addressing the poverty of stimulus problem.

Let me put this point another way. The issue of explanatory adequacy has been the 800 pound gorilla of grammatical inquiry and it has largely overshadowed the more standard benchmarks of theory evaluation. This is now changing for the following reason. As GB research has succeeded, a consensus has developed that P&P accounts answer Plato's problem in the domain of language. This consensus has served to cage the gorilla allowing other sorts of measures of success to drive theory construction, measures such as simplicity, elegance, parsimony, and naturalness. To put matters more starkly and tentatively than is warranted: given that P&P models solve Plato's problem the issue now becomes which of the various conceivable P&P models is best and this issue is resolved using conventional criteria of theory evaluation. In other words, once explanatory adequacy is bracketed, as happens when P&P proposals alone are considered, an opening is created for simplicity, elegance and naturalness to emerge as the critical measures of theoretical adequacy. This reorientation, however, prompts a question: how to concretize these heretofore subordinate evaluative notions in the specific research setting that currently obtains. It is here that minimalism aims to make a contribution. I turn to this next.

2 Economies in Theory Evaluation

To ask for the simplest most elegant theory based on the most natural sorts of principles often asks for very little. These notions are generally too obscure or subjective to have much practical purchase. To give them life we need to flesh out the problems against which theories are expected to measure up. Only then (and perhaps not even then) can we develop rough measures of theoretical beauty and parsimony. What then is the appropriate backdrop for linguistic theory? One way into this question is to recruit those facts about language, the "big facts," that any conceivable theory must address to be worthy of consideration. A second way is to develop simple parsimonious grammars that exploit "natural" thematically unified principles. Chomsky (1993) suggests ways of moving in both these directions.

MP exploits three kinds of considerations. First, it takes certain very general facts to be self evident and requires any theory of grammar to accommodate them. As noted in section 1, MP endorses the assumption that UG has a principles and parameters architecture. Other indubitable features of natural language (NL) include the following: (i) sentences are the basic linguistic units, (ii) sentences are pairings of sounds and meaning, (iii) there is no upper bound to the number of sentences in any given NL, (iv) sentences show displacement properties in the sense that expressions pronounced in one position are interpreted in another, and (v) sentences are composed of words organized into larger units with hierarchical structure, i.e. phrases. Together, these six facts serve as very general *minimal* conditions of adequacy on any theory of UG.

In addition MP deploys two types of economy considerations. The first type are the familiar methodological benchmarks such as simplicity and parsimony,

i.e. standard Ockham's razor sort of considerations: *ceteris paribus*, two primitive relations are worse than one, two levels are better than four, four modules are better than five, more is worse, fewer is better. Let's call such principles measures of *methodological economy*.

The reason that simplicity and parsimony are methodologically valuable is that they enhance the empirical exposure of one's underlying assumptions. To illustrate: if one can derive a body of data D using three assumptions, then D can be interpreted as lending empirical support to each of these assumptions. Each one carries part of the explanatory load and each is grounded to the degree that it is required to account for D. Note that if we reduce the required set of assumptions to two then this should, all things being equal, enhance the empirical support that D lends to each given that D is now spread over two assumptions rather than three.

Of course evaluation of alternatives is never this straightforward as things are never equal. There are trade offs that are hard to quantify between naturalness, parsimony and simplicity. However, the point remains that there are good epistemological reasons for adopting Ockham's strictures and trying to shave one's set of basic assumptions down to a minimum.

There is a second set of minimalist measures. Let's dub these *substantive economy*. Here a premium is placed on least effort notions as thematic sources for grammatical principles. The idea is that locality conditions and well-formedness filters are reflections of the fact that grammars are organized frugally to maximize resources. Short steps preclude long strides, derivations where fewer rules apply are preferred to those where more do, movement only applies when it must, no expressions occur idly in grammatical representations (i.e. full interpretation holds). These substantive economy notions generalize themes that have consistently arisen in grammatical research. Just think of the A-over-A condition (Chomsky 1964), the Principle of Minimal Distance (Rosenbaum 1970), the Superiority Condition (Chomsky 1973), the Minimality Condition (Rizzi 1990) and the Minimal Binding Requirement (Aoun and Li 1993a). It is natural to reconceptualize these in least effort terms. Minimalism proposes to conceptually unify all grammatical operations along these lines.⁶

These three sorts of considerations promote a specific research strategy: look for the simplest theory whose operations have a least effort flavor and that accommodate the six big facts about grammar noted above. This recommendation actually has considerable content. For example, the fact that sentences pair sounds and meanings and the fact that the number of sentences is essentially infinite requires both that grammars exist and that they interface with systems responsible for the articulatory/phonetic (AP) and conceptual/intentional (CI) properties that sentences display. Given this, there is a premium on grammatical principles that originate in this fact, e.g. if some sorts of grammatical objects are uninterpretable by the CI or AP interfaces, then phrase markers that contain these will be ill-formed unless these wayward objects are dispatched before interpretation. Given least effort criteria, the favored accounts will contain the simplest grammatical products that meet these interface requirements. This could mean the simplest to produce, in which case economy of derivational resources are key, or simplest to interpret, in which

case economy of representational resources (i.e. full interpretation notions) are highlighted.

Consider another set of questions minimalist considerations lead to. What are the basic primitives of the system; the basic relations, objects and operations? If phrases exist (i.e. (v) above) then a set of relations is provided if phrases are organized in roughly X' terms, as standardly assumed. In X' -theory, phrases have (at least) three parts – heads, complements and specifiers – and invoke two relations – head/complement and specifier/head. Given the obvious fact that NLS contain phrases, UG requires these objects and relations whatever else it needs. Therefore, parsimony counsels that at most these objects and relations should be part of UG. This implies, for example, that sentences be analyzed as types of phrases rather than as having an idiosyncratic structure. This is essentially the conclusion GB has already drawn. Labeling sentences as IPs, TPs and CPs embodies this consensus.⁷

The recognition that phrases are a minimally necessary part of any theory of grammar further suggests that we reexamine whether we need government among the inventory of basic grammatical relations. Methodological simplicity urges doing without this extra notion given that we already have two others. All things being equal, we should adopt government only if the X' -theoretic relations we already have prove empirically inadequate.⁸

The same reasoning extends to the inventory of rules in UG. It is self evident that natural languages manifest “displacement” in the sense that expressions in a sentence are heard in one position yet interpreted from another. Thus, grammars must have means of representing this. In GB, the basic means of accommodating this fact is via movement processes. MP requires that we treat this fact in the most parsimonious way possible. Grammars should therefore treat all instances of “displacement” in a unified manner. In GB, movement operations are dislinguished from construal processes. Construal rules are different in kind from movement rules and are used to analyze some instances of displacement. For example, control structures involve an expression pronounced in one place yet related to another position that contributes to its thematic interpretation. All things being equal, one set of rules is preferable. Thus, optimally, either movement is construal or construal is movement.

Assume, for sake of argument, that only movement rules exist. We can then ask how much of the GB theory of movement is motivated on minimalist grounds. Are traces, for example, conceptually required? In part perhaps, insofar as they simply model displacement (one of the big facts noted above) and provide a mechanism for coding the fact that expressions can be interpreted as if in positions distinct from the ones they overtly appear in. Does the simple fact of displacement motivate the GB view that traces are indexed categories without lexical contents, i.e. $[_{XP} e]_i$? Or does the existence of displacement phenomena suffice to ground the claim that traces are subject to special licensing conditions that do not apply to lexical items more generally? This is far less clear. Traces in GB are grammar internal constructs with very special requirements that regulate their distribution. Historically, the main motivation for traces was their role in constraining overgeneration in the context of a theory where movement was free, not in providing vehicles for interpretation.

The primary service traces and the conditions on them provided was to filter unwanted derivations that resulted from a grammar based on a rule like 'Move alpha'. Why assume that such entities exist, especially in the context of a minimalist theory in which it is assumed that movement is not free (as it is in GB) but only occurs if it must? Methodologically we should resist postulating traces as grammatical formatives unless strong empirical reasons force this conclusion. On conceptual grounds traces are of dubious standing.⁹

What could replace traces? Well, we independently need words and phrases. Why not assume that they are used by the grammar to accommodate displacement? In other words, assume that traces are not new kinds of expressions but they are copies of expressions that are already conceptually required.¹⁰ This seems simpler than postulating a novel construct if one's main goal is to accommodate displacement. In short, GB traces must earn their keep empirically and all things being equal a copy theory of traces is preferable.¹¹

What holds for traces holds for other grammar internal formatives as well; PRO, 0-operators and chains to name three more. It also brings into question the value of modules like the ECP, control theory and predication whose purpose is to monitor and regulate the distribution of these null (grammar internal) expressions. None of this means that the best theory of UG won't contain such entities or principles. However, minimalist reasoning suggests that they be adopted only if there is strong empirical motivation for doing so. On conceptual grounds, the burden of proof is on those who propose them. At the very least, minimalist scruples force us to reconsider the empirical basis of these constructs and to judge whether their empirical payoffs are worth the methodological price.

These sorts of abstract considerations can be easily amplified. The six facts MP takes as obvious make serious demands on GB style theories once issues of parsimony, naturalness and substantive elegance are taken as important measures of theory success. Both methodological and substantive economy lead to qualms about the adequacy of GB style theories. These in turn suggest grammars that have a different "look" from their GB precursors. Let's sample a few concrete MP arguments to see how these considerations are deployed.

3 Minimalism in Action

GB assumes that a grammar has four distinctive levels – DS, SS, PF and LF. PF and LF are conceptually necessary as they simply mirror the fact that sentences are pairings of sound and meaning. Thus, if there are grammatical levels at all, there will at least be a level that interfaces with the AP system and a level that contributes linguistic information to the CI systems. The other two levels, DS and SS, have a different status. If required at all, they are motivated on narrower empirical grounds.

Note that parsimony considerations favor the simpler two level theory unless there are good empirical reasons for postulating the more complex four level grammar. One minimalist project is to show that the two levels that are

conceptually required – LF and PF – are also empirically sufficient. Chomsky (1993) argues for this conclusion. Consider the reasoning. Let's begin with S-structure.

SS is useful in GB in at least three ways.

First, it is exploited in distinguishing languages like English from those like French with regard to verb movement and English from Chinese with regard to WH movement. English verbs do covertly (after SS) what French verbs do overtly (before SS) and Chinese WH words do covertly what English WH elements do overtly. SS, then, marks the divide between overt and covert syntax which, it appears, is useful in describing both the differences and commonalities among language specific grammars.

Second, SS is where case theory applies. Case is relevant both phonetically and semantically in GB. Phonetically, pronouns with different cases sound different; 'he' versus 'him'. Semantically, case marking is critical in that the Visibility Condition assumes that only case marked nominals are theta active at LF. Assigning case at SS meets both PF and LF requirements.

Third, the binding theory can apply at SS as well as LF in certain versions of GB to filter out unwanted derivations.

Chomsky (1993) argues that the first two problems can be accommodated without postulating an SS level if one assumes that movement is driven by a feature checking requirement and that features come in two flavors; weak and strong. Strong features must be checked prior to the point at which a derivation splits into separate LF and PF branches. Weak features, in contrast, can be discharged at LF. Using this technology, English verbs can be treated as bearing weak features checked at LF while French verbs bear strong features that require checking in overt syntax.

WH movement can be treated in a similar fashion.¹² So can case. For example, if accusative case is a weak feature, then it can be checked in covert syntax after the grammar branches. If nominative case is strong then overt movement is required to check it.¹³ Note that we have replaced the GB idiom of "assignment" with the terminology of "checking." Expressions enter syntactic derivations clothed with their features. These features get discharged/licensed via checking through the course of the derivation. As should be evident, this approach, in particular the combination of weak/strong features plus checking, eliminates the need for SS in theories of case as well as cross linguistic accounts of verb raising and WH movement. The new technology renders SS superfluous.¹⁴

This technology has a further interesting feature. It reflects the least effort themes of substantive economy. Movement is never gratuitous. It serves to license otherwise unacceptable items. Strong features are uninterpretable at either the PF (Chomsky 1993) or the LF interface (Chomsky 1995). Thus, these features would violate full interpretation were they to survive to the interface. Movement serves to "eliminate" such features by allowing them to be checked. In this sense, movement is recast in least effort terms as the way in which uninterpretable features get checked.

Movement in MP is a last resort operation in the sense of being illicit if it fails to result in some form of feature checking. Standard versions of MP

incorporate this idea in the principle that movement must be “greedy”: if A moves to K then either some feature of A or K is checked as a result. Greed, in effect, defines movement in the sense that non-greedy moves are simply not moves at all. Greed is the quintessential expression of substantive economy. Virtually all minimalist models exploit it in some form and thereby incorporate a least effort conception of grammatical operations.

The above illustrates the interplay of methodological and substantive economy considerations in MP. SS is suspect on methodological grounds. Thus, its utility must lie in its empirical virtues. But, the empirical motivations for SS fade if substantive economy notions are cast in terms of feature checking processes which reflect the least effort themes of MP. Interestingly, recasting grammatical processes as feature checking operations appears to be no less empirically adequate than prior approaches that exploited SS filters. Thus, there is less empirical motivation for SS than GB originally supposed. This supports the MP position, arrived at on general methodological grounds, that grammars have at most two levels, LF and PF.¹⁵

This conclusion invites further minimalist projects. Dumping SS requires reanalyzing all the phenomena that appear to exploit it. Chomsky’s (1993) reanalyses provide important first steps and indicate that the minimalist program has plausibility. However, there are additional phenomena whose standard accounts exploit SS in crucial ways, for example: parasitic gap licensing, predication, 0-operator licensing and island effects. Without SS, processes that have been described as holding at SS cannot be literally correct. Standard accounts of island effects, obligatory control phenomena, and 0-operator constructions such as ‘purpose’-clauses, tough constructions, relative clauses, and parasitic gaps rely on SS centered processes and so call for reanalysis. The succeeding chapters discuss possible ways of rethinking these phenomena in service of retaining the minimalist conclusion that UG has no SS level.

Chomsky (1993) also provides arguments for dispensing with D-structure. He argues that there can be no phrase marker that divides lexical insertion from the transformational component once ‘tough’-constructions are considered. The standard GB analysis of these constructions requires the interleaving of these two processes once mildly complex cases are considered. If so, DS, with the properties GB assigns it, cannot exist.

The argument Chomsky presents goes as follows. Chomsky, following standard GB practice, assumes that *John* is not base generated in the post verbal position of *please*, despite its thematic dependence on this post verbal position, but is directly inserted into the matrix subject position.¹⁶ This leaves ‘tough’-constructions with a property inconsistent with the GB notion of DS in that the syntactic subject is not base generated in a theta-position. (1b) indicates that the matrix subject has no thematic function as it can be filled by an expletive. The problem is how ‘John’ is inserted into the derivation while respecting the assumption that lexical insertion precedes movement.

- (1) a. John is easy to please
 b. It is easy to please John

Chomsky (1981) recognizes the problem posed by ‘tough’-constructions and weakens the DS thematic requirements in response. The idea was to allow simple NPs to be exempt from this requirement. Assume, for example, that ‘John’ can be inserted at DS without a theta-role so long as it gets one by LF. This allows lexical insertion to precede transformational rules in (1a) though at the cost of weakening the GB conception of DS as the place where grammatical functions and thematic roles meet. ‘John’ can be inserted in (1a) without a theta-role as it receives one via predication by LF. To repeat, this maneuver has two substantial costs. First, it weakens the DS requirement that insertion be exclusively to theta-positions. In addition, it complicates the grammar by adding a new kind of rule: predication. Predication fills the thematic gap that generating expressions in non-theta-positions opens up. This process, though distinct from the DS process of theta-marking, also serves to assign a theta-role to an expression. As such, a predication rule is a necessary complement to any attempt to retain the services of DS given the standard analysis of ‘tough’-constructions. These complications seem to allow GB to preserve DS as the border between lexical insertion and other kinds of transformations.

However, it does not work once slightly more complex samples of the ‘tough’-construction are considered. The amendment to the GB theory of DS aims to allow DS to mark the border between insertion and other transformations. However, it is possible to find subjects of ‘tough’-constructions that are transformationally formed. For example, a relative clause formed from a ‘tough’-construction.

(2) A man who is easy to please is easy to please

It appears that the subject here is formed via the application of rules such as relativization and ‘tough’ movement. Thus, it seems that the strict separation of movement from insertion transformations cannot be maintained. In sum, the GB version of DS as the level which is the output of lexical insertion, the input to the transformational component and “purely” represents thematic structure cannot be retained.¹⁷

This conclusion can lead to two different kinds of conclusions. One calls for reapportioning the tasks of DS to other parts of the grammar. Recall that it is uncontested that NPs are infinite. Thus, any theory of UG must have a mechanism for recursion to be adequate. In GB, recursion is lodged in DS, in the phrase structure rules of the grammar. MP substitutes generalized transformations for these phrase structure rules and makes them the recursive engine of the grammar.

Similarly, Chomsky (1995) relocates parts of theta theory to the transformational component. The GB hypothesis that DS is the domain in which thematic functions get defined is analyzed in MP as a condition on phrase formation. It is assumed that only trivial chains can assign or receive theta-roles, i.e. technically, theta-roles are assigned under Merge not Move. This has the effect of leaving theta-role assignment sensitive to base (i.e. pre-movement) configurations but without adverting to any level akin to DS. In short, as was the case with SS, Chomsky shows that garnering the services of DS conditions does not

require postulating a DS level. Once again, the MP conclusion is that DS, like SS, is not required and so can be dispensed with as it is without empirical or conceptual warrant.

There is a second kind of conclusion one can draw. One can begin to question whether the conditions generally assumed to apply are really required at all. Note that this is different from assuming that the generalizations are correct but do not necessitate the technology heretofore deployed. This sort of conclusion has a more radical tinge in that it asks that the generalizations that have been taken as established be reevaluated. For example, Chomsky (1995) retains a central feature of DS in the idea that trivial chains are the only proper vehicles for theta role assignment or reception. His reasoning relies on analyzing theta-roles as very different from morphological features such as case and phi-feature agreement (c.f. chapter 2 for discussion). How good are the empirical and conceptual arguments for this distinction? I argue that there are advantages in rejecting it. The payoff is a more general conception of movement and a simpler picture of universal grammar. More specifically, the logic of MP invites the following line of investigation.

Jettisoning the vestiges of the GB conception of DS has two potentially positive consequences. First, it allows for the elimination of PRO from the inventory of empty categories. Second, it allows for the elimination of the PRO module from UG. Each consequence is attractive on methodological grounds if empirically sustainable. The conceptual superiority of fewer levels extends to modules too; one less is better than one more! As for PRO, it is a theory internal abstract entity whose worth must be supported on empirical grounds. All things being equal, it is no better to have grammar internal abstract entities than it is to have grammar internal interfaces.

This line of inquiry has another consequence. It suggests an extremely general approach to movement, one in which it is completely unrestricted. GB permits movement from theta to non-theta-positions and from non-theta to non-theta-positions. However, it forbids movement from one theta-position to another or from a non-theta-position to a theta-position. These restrictions follow from the role that DS plays in theta-role assignment in a GB theory and from the analogue of this restriction concerning trivial chains that Chomsky's version of MP incorporates.¹⁸ All things being equal, however, these restrictions are undesirable as they stand in the way of the most general approach to movement; anything can move from anywhere to anywhere. Any restrictions require strong empirical support as, *ceteris paribus*, they complicate the operations of MERGE and MOVE.

This general conception of movement is in turn part of a larger project: to use movement as the primary vehicle for establishing grammatical coupling, e.g. for coding internominal dependencies intrasententially. Assume that core grammar only contains the simple operations of MERGE, COPY and DELETE.¹⁹ Movement, let us assume, is COPY+MERGE.²⁰ The applications of MERGE and COPY result in various kinds of feature checking. I argue in the following chapters that almost all core grammatical relations can be analyzed in terms of these processes if they are completely generalized, in particular, if one drops the thematic restrictions on movement. Expressions check and gain features

via movement and they establish anaphoric relations by movement. If this is correct, there is no control module or binding module and grammars contain no analogues of rules of construal in the GB sense.

One could go further still. The possibility arises that strictly speaking UG is *not* composed of separate modules with specific organizing principles and constraints. Rather, the main grammatical operations – MERGE, COPY, DELETE – apply to all features. Different applications of MERGE and MOVE (COPY+MERGE) may check different features – case, agreement or theta-features for example – but the way in which all these features are checked is essentially the same. If this is correct, the GB picture of interacting modules placing different kinds of well formedness conditions on phrase markers, exploiting different domains of application and different sorts of rules is replaced by a theory in which grammars check features by merging and copying expressions in the same way for every feature. As the operations involved are the same, the locality domains and restrictions should hold indifferently across the various types of checking operations. Just as modules in GB allow the elimination of construction specific rules, the proposal here is that there are no morpheme specific processes either; no grammatical rules that specifically target PROs or anaphors or pronouns.

This is the picture. I believe that this conception is implicit in one version of MP. It arises from considering how various aspects of the program have been articulated and considering to what extent GB fits in both with these details and with the larger methodological and substantive economy issues noted earlier. Let me expand on this a bit.

For concreteness consider the GB Binding module. It is suspect in several ways from a minimalist perspective. First, it exploits different kinds of rules from those found elsewhere in the grammar. Rules of construal are indexing procedures additional to the one that ‘Move alpha’ already embodies. GB thereby contains two kinds of indexing operations. Parsimony urges UG to make do with just one, all things being equal.²¹ So, either we eliminate movement processes or construal rules.

The reduction of movement to construal was investigated in earlier GB work by Koster among others.²² This is achieved by treating traces as lexical items that are base generated. Subsequent rules of construal provide the relevant relations via indexing algorithms of various kinds. However, this requires treating traces as *lexical* expressions with distinctive needs. This is a high price to pay if one aims to eliminate morpheme specific processes. Moreover, the advantage of so recasting movement is quite unclear. These approaches generally postulate two kinds of indexing processes, those subject to the locality conditions characteristic of movement and those not subject to these restrictions. We thus end up with the same reduplication of procedures embodied in the earlier GB accounts (c.f. Chomsky 1981).

Note that the presence of displacement phenomena is uncontroversial. The only question is what rules the grammar exploits in accommodating it. The standard assumption has been that movement processes in some form are ineliminable. Thus, if there is to be a reduction it plausibly goes in the direction of analysing construal as movement rather than the reverse.

Some GB work has examined this alternative in the domain of anaphora. Chomsky (1986b), building on work of Lebeaux (1983), proposes that ‘self’-anaphors move covertly at LF to positions close to their antecedents. Anaphor movement does not replace Principle A of the binding theory in Chomsky (1986b). However, once operative, Principle A appears to be redundant in that the antecedence relation that it establishes can just as easily be treated as the by-product of the movement alone. This, in fact, has been tacitly concluded by the field. GB, then, began the move towards reducing Binding Principle A to (A-) movement, at least in cases of local anaphora.

Observe that this “reduction” serves to make unnecessary the locality restrictions on reflexives induced by the binding theoretic notion ‘domain’. From a minimalist perspective, this is a very positive result. Consider why.

Binding domains are defined in terms of government.

- (3) D is a domain for β iff D is the least complete functional complex containing β and a governor for β

MP considers the government relation methodologically suspect on grounds of parsimony. The world would be a better more elegant place if government were not a primitive relation required by UG. Reducing anaphora to movement allows Principle A of the binding theory to be dispensed with and thereby removes one reason for retaining government as a primitive grammatical relation.

In short, reducing anaphora to movement serves to simplify things in several respects. First, it would be a step towards eliminating construal rules as separate grammatical operations thereby allowing the inventory of rule types to be reduced. Second, it would be a step towards removing ‘domain’ as a theoretical construct. As the locality properties of anaphors would reduce to those of movement, there would be no need for binding domains to circumscribe their distribution. Last of all, removing domains from UG also aids in removing government as a basic primitive relation of the theory. As this notion is independently suspect, any move towards its elimination is welcome.

Reducing Principle A to movement, should it prove successful, immediately adds another item to the research agenda: how to eliminate Principle B. Assume for the sake of argument that the distribution of anaphors can indeed be reduced to the theory of movement. This would then place Principle B in a very odd light. Why should there be a principle of grammar whose main concern is the distribution of pronouns? What makes them so special? Moreover, even if the grammar does care about pronouns why should it devise *sui generis* relations (i.e. government), operations (i.e. construal rules) and locality restrictions (i.e. domains) to determine their distribution? Without anaphors, domains serve exclusively as vehicles to specify the distribution of pronouns. This further enhances the *ad hoc* status of the notion and renders it (and the notion of government that defines it) yet more suspect.

The problem is not just a theoretical one. Once the distribution of anaphors is reduced to movement, an empirical puzzle emerges: why is it that pronouns

and anaphors are generally in complementary distribution? The GB approach to binding accounts for this observation, which, though not perfect, is surprisingly robust.

The near complementary distribution of (local) anaphors and pronouns follows in GB given the nature of principles A and B of the binding theory.

- (4) A: an anaphor must be bound in its domain
 B: a pronoun must be free (i.e. not bound) in its domain

Principles A and B make opposite demands on anaphors and pronouns. If one assumes that domains for anaphors/pronouns are defined as in (3) this results in anaphors being licensed by Principle A if and only if (bound) pronouns are forbidden by Principle B. Thus, one important empirical consequence of the GB binding theory is that anaphors and pronouns should be in complementary distribution. This seems to be a largely correct description of the facts.²³

However, once one abandons Principle A and makes the distribution of anaphors the province of the theory of movement (as has been standard practice since Chomsky (1986b)) the above GB account evaporates. There is no obvious reason why (near) complementarity should exist between pronouns and anaphors if they are regulated by entirely different modules of the grammar. As should be evident, there are good reasons for finding the GB Binding theory wanting. The MP concerns simply bring its inadequacies into sharper focus.

The foregoing is meant to illustrate a general feature of MP. Once one starts to pull on part of the theory, the economy considerations that drive the program quickly lead to a general unravelling of the theory as a whole. It is in this sense that MP is potentially so far reaching and radical.²⁴ No part of the theory is very remote from any other part and changes in one domain naturally lead to questions about others when global evaluations of simplicity, parsimony and elegance drive theory choice.

A second important point to note is that methodological concerns such as parsimony get their bite when pairs of possibilities are played off against one another. Theories are neither simple nor complex, neither parsimonious nor profligate simpliciter. They must be as complex and intricate as required. Thus, when accounts are considered singly evaluations of methodological economy are moot. It is only in the context of theory *comparison* that such notions find a foothold. As such, it always pays to have a competing companion account for purposes of comparison. Absent this, methodological considerations quickly lose their grip and utility.

GB accounts admirably fit the role of straight man to the minimalist kibbitzer. One way of fruitfully launching a minimalist research program is to simplify, naturalize and economize earlier GB accounts. These are always good places to begin and provide solid benchmarks against which to measure putative progress. I adopt this comparative strategy in what follows.

4 Some Specific Principles

The earlier sections have tried to instill a feel for the global ambitions of MP by illustrating the types of principles assumed and how they are invoked to generate questions and projects. Now it is time to outline some specific grammatical principles that will serve as fixed points in the chapters that follow. These (hopefully) reflect the larger themes outlined above. However, they are more specific and constitute particular ways of concretizing these minimalist sentiments. There are surely other ways of pursuing the broad outlines of the program. Some of the analyses that follow argue for the specific principles outlined below in being required if these analyses are to succeed. More often, however, the chosen implementation is just one of many ways of setting the stage and other arrangements could serve as well. For concreteness, I list the relevant principles.

4.1 *There are only Two Grammatical Levels, LF and PF*

For all practical purposes, only LF has grammatical standing as PF is too unstructured. Thus, whatever filters apply, do so at LF. This does not mean to say that there are no bare output conditions imposed by PF. For example, the Linear Correspondence Axiom (LCA), the algorithm that takes a hierarchical structure and produces a linearization of its constituents, is a plausible requirement on phrase markers imposed by the PF requirement that expressions be pronounced in a particular serial order. This follows Chomsky's (1995) interpretation of the LCA.²⁵ Chomsky notes that it is natural to think of linearization as a PF requirement given that phonological operations are sensitive to the linear properties of strings. As such, the LCA can be seen as part of the contribution the grammar makes to the interpretive requirements of the PF interface.

In the best case, filters reflect "bare output conditions," i.e. conditions imposed on LF (or PF) from the fact that it is a level that interfaces with the conceptual/intentional systems (CI) (or the articulatory/phonetic (AP) systems) whose intrinsic properties impose conditions on LF (or PF) phrase markers. The above interpretation of the LCA is one example of this. Consider an example on the LF side.

It is reasonable to accept Full Interpretation (FI) as a condition imposed by CI on LF phrase markers. FI requires all features that pass across the interface to receive an interpretation. On this conception, FI is a bare output condition that filters out sentences containing expressions that have unchecked uninterpretable features, e.g. unchecked case features. In effect, on this view, the case filter simply reflects the requirements of the CI interface. It is not a specifically language internal requirement.²⁶

There are other LF filters that plausibly have similar interpretations, e.g. the recoverability of deletion, parts of the theta-criterion, viz. the requirement that

theta-roles of a predicate be expressed and the requirement that all DPs have theta-roles. On the other hand there are many requirements that fit less well, e.g. the various locality conditions on anaphors and pronouns imposed by standard versions of the binding theory and the requirement that DPs have at most one theta-role. There seems nothing inherent in the notion pronoun or anaphor that brings with it the requirement that it meet its antecedence conditions within fixed domains nor with the notion argument that restricts it to having but one role. This contrasts with the requirement that an anaphor have an antecedent or that a pronoun can be bound. These conditions do plausibly follow from the inherent interpretive properties of such expressions.

4.2 *There is a Fundamental Distinction Between Case and Theta Theory*

This is executed in various ways within MP. For example, Chomsky (1995) takes case to be a feature checked in functional domains. Thus, for example, an expression bearing case can check this feature by moving to the Spec of a phrase headed by a functional expression that bears such a feature, e.g. the Spec of TP for nominative case and the Spec of AgrOP or the Spec of vP for accusative case. Theta-roles, in contrast, are *not* features and are (typically) assigned within lexical domains, i.e. phrases headed by lexical expressions, to trivial chains and by trivial chains.²⁷

An interesting consequence of this bifurcation is that it implies some version of the predicate internal subject hypothesis, the idea that theta-marked subjects get their roles within the predicate phrase, (e.g. VP) rather than in Spec IP.²⁸ In the context of MP this has a further attractive implication. MP aims to pare down the required grammatical operations to a minimum. The predicate-internal subject hypothesis allows for the elimination of operations like “predication” as follows. If subjects are generated in Spec IP then they must be assigned theta-roles via a rule of predication. This rule is different from standard forms of theta marking in which roles are simply assigned by heads within their local phrasal domains. If rules like predication are dispensed with then subjects must receive their roles in some other way. Thematically marking subjects within lexical shells and raising them to Spec IP dispenses with the need for predication in such cases.

In the chapters to come, I follow Chomsky partway. I assume that the domains of case and theta theory are distinct; the former being a relation between a D/NP (or its features) and a functional head (e.g. the Tense or Agreement), the latter (typically) a relation with a lexical head (e.g. V or N). However, I drop the other two assumptions.

In particular, I treat theta-roles as features in at least one important sense: like all other features they are able to license movement. This contrasts with Chomsky (1995, 1998) where it is assumed that only morphological feature checking can license movement. Concretely, I assume that theta-roles are assigned by heads (mostly verbs are examined) to D/NPs that merge with them. This allows any instance of MERGE to be a potential theta discharge

configuration. Thus, movement can result in theta-marking and is not limited to non-trivial chains.²⁹

4.3 Features are Checked in Configurations Licensed by Phrasal Structure

As noted in section 2, the theory of phrase structure requires at least two kinds of relations – head/complement and specifier/head. Chomsky (1995) assumes that the former is the most basic configuration with Spec/head being the residue of the head/complement configuration. Thus, spec/head relations encompass the standard cases of spec/head configurations as well as head/head relations formed via head movement. Both ‘ZP’ and ‘Z⁰’ in (5) are in spec/head relations with ‘X⁰’.

- (5) a. [ZP [X⁰ YP]]
 b. [[Z⁰ [X⁰]] YP]

Case features are checked in spec/head relations.³⁰ Internal theta roles are checked in head/complement configurations while non-internal roles are checked in spec/head structures.

The assumptions in section 4.2 and section 4.3 constitute one important difference between GB and MP. GB aimed to unify the configurations of case and theta assignment under government. MP questions this goal. Rather, it unifies what it sees as disparate instances of case assignment by treating all instances as the reflex of a spec/head relation between a nominal and a functional head. Given that theta-roles are assigned in lexical domains, this requires that nominals move to check case in MP. The movement is necessarily to some non-complement position, i.e. it results in a spec/head configuration.

The conclusion that case is checked in derived positions is one of the most interesting claims in MP and it is worth pausing to consider the theoretical and empirical reasons behind it. The GB theory of case tries to unify case assignment under the government relation. Government is defined so that verbs govern their canonical objects, inflections govern subjects and ECM verbs govern the subjects of their sentential IP or small clause complements, c.f. (6).

- (6) a. [_{VP} V D/NP]
 b. [_{IP} D/NP [_{I'} Infl . . .]]
 c. [_{VP} V [_{IP/SC} D/NP . . .]]

From a minimalist perspective, this unification of the circumstances of case assignment seems rather contrived. (6a) is an instance of a head/complement relation, (6b) a spec/head relation and (6c) neither of these. The last is clearly incompatible with a phrasal source for basic grammatical relations. However, even the difference between (6a) and (6b) suggests that the GB approach to case is not conceptually unified.

The lack of structural homogeneity in the three instances of case assignment suggest that a revision is in order. (6) indicates that case cannot be unified under the head/complement relation. The only alternative (given MP) is the spec/head relation. In short, the MP theory of case requires taking (6b) as the canonical case configuration. The nominals in (6a,c) must move to configurations analogous to (6b). Chomsky (1993) takes the target of this movement to be Spec AgrO while Chomsky (1995) assumes it is the outer Spec of vP. In any case, this means that objects and ECM subjects have moved out of their VPs or clauses to higher Spec positions to check case.

This reasoning gets interesting support from binding data. Lasnik and Saito (1993), resurrecting earlier observations of Postal (1974), argue that ECM subjects appear to have wider scopes than their overt positions would support. For example, the ECM subject in (7a) appears able to bind the reciprocal in the matrix adjunct, in contrast to the embedded subject in (7b).

- (7) a. The DA proved [the defendants_i guilty] during each other's_i trials
 b. *The DA proved that [the defendants_i were guilty] during each other's_i trials

The contrast follows on the assumption that 'the defendants' moves (perhaps at LF) to the matrix clause to check (exceptional) accusative case while it remains in the embedded clause when checking nominative. On this assumption, 'each other' is in the scope of the ECM subject but outside the scope of the nominative.

4.4 *Movement is Greedy*

Earlier sections outlined show this fits in with the general precepts of MP. For present purposes I interpret this requirement as mandating that movement is only licensed if the resulting structure allows for the checking of some feature. In current parlance, I assume that greed is "enlightened."³¹

- (8) MOVE allows A to target K only if a feature of A or K is checked by the operation

Other versions of greed require a "moved" expression to check one of its own features. (8) allows either the moved expression A or an expression in the target K that comes into relation with A to check some feature. This is now a common way of understanding greed.

4.5 *Movement is Actually the Combination of Copy and Merge*

This view is proposed in Chomsky (1993) and defended extensively in Nunes (1995). Chomsky indicates how making this assumption allows one to dispense with certain binding arguments for SS. However, the issue is more general.

Analyzing movement as a complex of two simpler operations – COPY and MERGE – allows one to dispense with traces as primitive grammatical formatives, as noted above. Lexical copies can replace traces for marking the positions from which movement takes place.

Moreover, viewing movement as composed of Copy and Merge may allow one to account for other postulated properties of MP grammars. For example, Chomsky (1995) proposes that Merge is a more economical operation than Move in order to account for the unacceptability of sentences like (9).

- (9) a. *There seems a man to be here
 b. There seems to be a man here

The explanation proceeds as follows. On building the phrase marker for (9) one starts with a lexical array of items and progressively builds the sentence by Merging and Moving expressions. Consider the point at which (10) has been formed.

- (10) Infl⁰be [_{SC} a man here]

At this point, it is possible to move *a man* to the Spec of Infl or to Merge *there* from the array into this same position. If Merge is less costly than Move then the latter option is preferred. This allows for the derivation of (9b) but not (9a).

This proposal requires that Move be more costly than Merge. Treating Move as the combination of Copy and Merge has this as a trivial consequence given that the operations that underlie Move properly include Merge. Hence, the number of operations required to extend the derivation using Merge is less than those using Move.

So treating Move has other desirable consequences. Chomsky (1998) uses it to account for reconstruction effects in the grammar.³² It further permits the formulation of a simple generalization: all features are checked via MERGE. This generalization is methodologically desirable. Consider why.

If MOVE is a totally different operation from MERGE then feature checking could be licensed either by merging two categories or by moving a category from one position to another. Economy considerations dictate that there be only one way of checking features. Given that MERGE is a virtually conceptually necessary operation (it follows from section 4.5 above which states that atoms (i.e. words) can combine to form more complex structure) it would seem that MOVE should be related to it rather than postulated as an entirely distinct kind of process. Resolving MOVE into the pair of operations COPY and MERGE does just that. Thus, the reason that both MERGE and MOVE lead to feature checking is that both involve a common MERGE component.

Note one last point. This suggests that what makes MOVE expensive is the fact that it involves copies. Making copies costs! Copies are tolerated only if they promote convergence by eliminating uninterpretable features. This reasoning rationalizes the link between MOVE and greed which is part of the standard MP package of assumptions.

4.6 *Overt Instances of MERGE and MOVE Adhere to the Extension Condition (Chomsky 1993)*

The Extension Condition is the requirement that grammatical operations enlarge the tree in the sense that the phrase marker that results from the operation contain as a sub-constituent the phrase marker that was input to the operation. (11) illustrates the condition.

- (11) a. [A .. [B ... [C ...]]]
 b. [D [A ... [B ... [C ...]]]]
 c. [A .. [B ... [C . D . .]]]

Assume that (11a) is the input to the relevant operation. Then “adding” *D* as in (11b) obeys extension in that the phrase marker in (11a) remains as a proper subpart of (11b). (11c) is not a licit phrase marker for it “thickens” rather than extends the phrase marker. Note that the phrase marker (11a) is no longer a constituent of (11c).

A consequence of the Extension Condition is that overt syntax adheres to the strict cycle and also prohibits lowering operations from applying in overt syntax. Chomsky (1998) suggests (following Epstein (1999), Epstein, Groat, Kawashima and Kitahara (1998), and Kawashima and Kitahara (1996)) that the restriction of the Extension Condition to overt movement might be reduced to the LCA by assuming that command is a derivationally determined relation. I here assume that extension applies to overt syntactic operations whether or not the Extension Condition can be derived from more general principles. I follow Nunes (1995) in assuming that the Extension Condition also applies to adjunction. These issues are discussed in further detail in chapter 3.

All of these assumptions are standard though some of the implications I draw from them are not. They are more or less drawn from Chomsky (1997). However, there are a host of standard assumptions that are rejected in what follows. I list some of them below and return to justify these departures from the current consensus in the relevant chapters.

4.7 *Theta-Roles Are Not Features*

This assumption is rejected in order to reduce parts of the theory of control to the theory of movement. The details are outlined in chapter 2. Note, the assumption that theta-roles are features does not imply that they are identical in all respects to morphological features like case. I continue to crucially assume that the domains within which case features and theta-features operate are distinct. However, for the purposes of MOVE and greed theta-roles are no less able to license operations than case features are.

4.8 *The Binding Theory Applies at LF*

Section 3 has argued that there are minimalist reasons for questioning the status of the Binding Theory in MP. This is investigated more fully in chapter 7.

4.9 *The Theta-Criterion Does Not Hold*

This is crucial for the analyses in chapters 2 and 3. For example, chapter 2 argues that control can be reduced to movement. A necessary assumption for this sort of analysis to work is that a DP be permitted to move into more than one theta-position. This, however, requires dispensing with the theta-criterion which is intended to prevent such movement.

4.10 *Movement is Actually Attraction*

This is proposed in Chomsky (1998) and (1999). Chapter 4 argues against the Attract based conception of displacement.

5 Conclusion

The proof of the pudding is in the eating. The ingredients for the MP pudding I propose to make here have been outlined above. I have also attempted to outline the broader MP concerns and show how they interact to motivate various moves in the minimalist research program. This prolegomenon will hopefully serve to guide the reader through the thickets of the remaining arguments.

One last point. Let me reiterate what I said at the outset. Minimalism is a program, not a theory. The program, if successful, will prompt the creation of various minimalist models of grammar each of which gains inspiration from the sorts of considerations outlined above. These models will differ in (at least) (i) what they take the broader issues that motivate minimalism to be (ii) how they weight these broader concerns (iii) how they implement them in particular analyses and (iv) other things I haven't thought of. As such, when in what follows I lapse into talking as if the particular combination of assumptions I am exploring is the "one and only true path," I implore the reader to have a good chuckle at my expense, forgive the slip, and move on.

Notes

- 1 The discussion here stresses the epistemological underpinnings of minimalism. There are alternative minimalist visions that articulate the program from the perspective of a metaphysics of complexity. What I say below is compatible with this

- view but does not presuppose it. For discussion of this alternative see, for example, Uriagereka (1999) and Martin and Uriagereka (2000).
- 2 It appears that many doubt Chomsky's sanity as they appear to believe that he and his followers deny this trivial point. See, for example, Bates and Elman (1996).
 - 3 This has recently been challenged in Cowie 1999. For a critical review of Cowie that I personally find completely decisive see Crain and Pietroski (1999).
 - 4 This is a vastly simplified picture. However, little is gained in going into the complexities here. For some discussion see Chomsky (1986b), Dresher (1998), Dresher and Kaye (1990), Hornstein and Lightfoot (1981), and Lightfoot (1982, 1999).
 - 5 For discussion see Chomsky (1983).
 - 6 The history of physics demonstrates the power of pursuing thematically consistent theories. The search for inverse square laws gave way to the pursuit of conservation principles which in turn was replaced by an interest in field equations which itself yielded to theories based on symmetry principles. For discussion of how these themes drove research see Weinberg (1988).
 - 7 This conclusion was based on work by Stowell (1981), Chomsky (1986), and Koopman (1984) among others.
 - 8 Chomsky (1995) argues that X' properties of phrases follow from very modest conceptual assumptions. If his Bare Phrase Structure hypothesis is correct it further strengthens the point made here that X' relations have conceptual priority over notions like government.
 - 9 Nunes (1995) forcefully makes this point. Chomsky (1998) has recently endorsed the same conclusion.
 - 10 Chapter 6 argues that copies are conceptually required objects within grammars that postulate a distinction between the lexicon and the computational system. If so, copies are conceptually costless. See chapter 6 for discussion.
 - 11 In GB, traces are part of the definition of Move. This suggests that within MP, movement be defined as involving Copy as a suboperation. In other words, there is a very short conceptual distance within minimalism between eliminating traces as undesired grammatical entities and adopting the copy theory of movement. See below for further discussion of the copy theory of movement.
 - 12 Though this is not the route taken in Chomsky (1995: ch. 3). He follows Watanabe (1992) and Aoun and Li (1993b) in assuming that WH movement always takes place in overt syntax even in languages like Japanese and Chinese.
 - 13 Chomsky (1995) assumes that both nominative and accusative are weak but that another feature, the EPP feature, is strong and forces nominatives to raise overtly. For present purposes these details are irrelevant.
 - 14 There is nothing particularly "minimalist" about feature checking, nor is it especially interesting to check rather than assign features. What is interesting is Chomsky's (1993) claim that adding features to standard GB operations eliminates the need for SS conditions. In other words, what we find is that a rather trivial technical change of implementation eliminates one of the strongest arguments for what we took to be one of its striking features, viz. the existence of a level of S-structure.
 - 15 See Chomsky (1995: ch. 3) for a discussion of how to eliminate the Binding Theory from applying at SS.
 - 16 See Chomsky (1981) for the standard GB analysis of such constructions.
 - 17 To be honest, it is not clear why 'tough'-constructions are required to make this argument. It appears that the nub of the argument relies on the proper treatment of relativization. The assumption seems to be that relative clauses are formed by movement and until formed, they cannot be merged. However, if so, then a sentence

like (i) could just as easily have been used to make the case. The detour through ‘tough’-constructions is irrelevant.

(i) John met the man who I like

Proponents of GB, however, might well resist the conclusion that movement is required for having an NP sufficient for insertion. After all, phrase marker rules were developed to generate relative structures. What movement then does is relate the head to a theta-position internal to the relative. It is unclear why this is illicit.

This said, ‘tough’-movement structures do introduce serious complications into the GB view of DS, as noted in the text, and these suffice by themselves to cast doubt on its value. See chapter 3 for a possible reanalysis of these constructions.

- 18 Parts of the theta-criterion are plausibly part of the Bare Output Conditions that characterize the interpretive properties of the CI interface. However, one feature is clearly grammatical: the requirement that DPs and theta roles be bi-uniquely related. There is little conceptual reason for prohibiting a DP from having more than a single theta-role. See chapters 2 and 6 for more discussion.
- 19 See chapter 6 for some discussion of DELETE.
- 20 This view of movement is first examined in detail in Nunes (1995). It is adopted and elaborated below.
- 21 I do not intend to treat indexing as a “real” process. Rather it stands surrogate for real relations among expressions. GB conceives of at least two distinct ways of doing this; via movement and via rules of construal.
- 22 See Brody (1995) for a version of Koster’s (1978) program with a minimalist accent.
- 23 For a recent discussion of this complementarity see Safir (1997).
- 24 It also indicates that GB must have a rather interesting theoretical structure if pulling on one part leads so quickly to reconsideration of so many other aspects of the theory.
- 25 Kayne (1994) first proposes the LCA. However, he takes it as a condition on all phrase markers, including those at LF. Chomsky (1995) notes that Kayne’s arguments do not accomplish what is desired nor are they necessary. In chapter 3, I suggest a reinterpretation of Kayne’s original idea so that something analogous to the LCA holds at LF.
- 26 Note, the requirement that expressions bear morphological features *is* an internal requirement of the language system! Why this is required is unclear. However, there is strong evidence that it is. Note, moreover, that case is linked to the fact that displacement exists in NLS. Just what the conceptual relation is between these two facts, however, is still unclear. Chapter 6 offers an account for why uninterpretable features might be part of an optimally designed system.
- 27 I say *typically* as the external theta-role is assigned in the Spec of vP in Chomsky (1995). This small *v* has a standing in between the lexical and the functional.
- 28 See Kuroda (1988), Koopman and Sportiche (1991).
- 29 This is also examined by Bosković (1994), Boskovic and Takahashi (1998), and Lasnik (1995b).
- 30 At least in languages like English. There is nothing in MP that forbids checking case in head/complement configurations and there is some research that indicates that this indeed happens in some languages. In what follows, I put this possibility to one side.
- 31 C.f. Lasnik (1995b).
- 32 See Aoun, Choueiri and Hornstein (1999), Fox (1999), and Kim (1998) among others for other examples that tie reconstruction to movement.