Merge over Move and the empirical force of economy in Minimalism

Alexandra Motut
University of Toronto

This paper presents an overview of the origins and history of economy principles in Minimalist syntax, and investigates the empirical justification of one particular economy principle, Merge over Move (MOM). It is argued that the principle is not well-supported empirically or conceptually, and that the primary examples that motivated the introduction of the principle can be accounted with other, alternative analyses. The residual phenomenon of subject control into adjuncts in English is then addressed, an empirical domain where Hornstein (2001) has argued MOM derives the correct empirical results. This specific critique of the MOM principle is embedded in a broader discussion of economy principles and the ad hoc nature of their proposal in much previous work. It is argued that a “unified” economy, operating broadly in the grammar and subsuming more specific principles, remains elusive. Finally, some potential theoretical consequences of abandoning MOM are examined.

1. Economy in Syntax: What is it and why care about it?

While some worry about the state of the world’s economy, I have been spending a great deal of time worrying about another kind of economy: economy as a metric of the parsimony of syntactic models in theoretical linguistics. Economy appears very early in theories of syntax, but only recently has become a more prominent focus of attention. Even now, to many working in syntactic theory, economy is not a primary concern. It is possible to build analyses that attain several levels of adequacy without ever having to venture a stake in a particular formulation of economy. But concerns about economy frequently creep into the theories, and can even be said to underlie the most recent of them (i.e. Minimalism), and thus it is fruitful to examine economy in a critical light to see where it leads us. Economy considerations can be used as justification for the choice of one derivation over another, or as an ad hoc explanation for an empirical contrast that otherwise would go unexplained. A key aspect of my arguments in this paper will be that there is a kind of imprecision in how economy is applied to syntactic models. Economy has many different faces, and I will argue that these several faces constitute a theoretical as well as a methodological problem that justifies investigation.

First of all, what do I mean by economy? In syntax, a grammar, or a model of a grammar, is economical if it is doing the least work possible in order to generate grammatical forms, and rule out ungrammatical forms. Most formulations of economy rely on certain axiomatic assumptions as to

* This work is essentially a version of my Master’s thesis, or M.A. Forum Paper, from the University of Toronto. As with all work of this kind, much further work remains to be done, and thus this is still very much a ‘working’ paper. I am extremely grateful to a number of individuals who helped make the completion of this paper possible, including Elizabeth Cowper, my supervisor; Elan Dresher, my second reader; Derek Denis; and, the many individuals with whom I had formal and informal conversations on the topic. Thanks also to audiences at the Syntax Project Research
what the minimal components of grammar must be. Within Minimalism, for example, these components are two interfaces, LF (logical form) and PF (phonological form), and a lexicon. In fact, Collins (1997: 2) goes so far as to state that, “any theory of syntax must contain a lexicon, a PF interface, and an LF interface” (italics mine). A model, or a grammar, is economy-based if it excludes ungrammatical forms via some economy principles or conditions. Economy conditions rule out ungrammatical forms, or particular derivations, based on their being too “costly,” or such principles choose between alternative derivations\(^2\) based on what is less expensive, computationally. In other words, if the model lives up to the motto, *If there’s a simpler way to do it, do it that way,* then it abides by economical principles, or economic principles can be said to be at work in the system. Of course, this raises the question of what it means to be (computationally or otherwise) “less expensive” or “less costly” in the syntax. It is also true that economy can be defined over different components of the model (i.e. not only the syntax). Here I talk mainly about economy of the (narrow) syntax, but it is possible to define economy as ranging over more than simply the syntactic component of a grammar. Brody (1995) argues for a kind of economy that takes LF within its range, rather than the kind of economy that takes LF as an output level. The Minimalist Program (MP), on the other hand, advocates the kind of economy that starts with specific assumptions about the input and output levels (i.e., the lexicon, LF and PF), and then economy basically ranges over the path a derivation takes from the lexicon to the output levels (the ‘narrow syntax’).

Economy has its uses: primarily, it is invoked to rule out ungrammatical sentences, though as we will see, the cases in which it does so, and the forms it takes when doing so, are inconsistent. Defining the correct measure of economy has proven difficult: on the surface, it is not obvious that the grammar functions economically. Functionally, many grammatical forms may express similar or identical meanings. It is also not immediately apparent what kind of economy is empirically valid. We may also distinguish methodological economy from substantive economy (the distinction described in Hornstein et al 2005: 8); the former pertains to how we go about modeling the grammar of the human language faculty, whether we do so in the spirit of Occam’s Razor, and the latter concerns verifiable empirical evidence that UG actually operates based on economy conditions. Besides the fact that in certain cases economy principles seem to be necessary to rule out unwanted derivations, there is also a second motivation for economy in the syntax: we like the idea of parsimony in a theory, and while some extend this parsimony to UG itself, and others limit it mainly to methodological concerns, in general the idea of a simpler system is more aesthetically appealing.

I will start with a review of the different ways in which economy has been instantiated in syntactic models. First, depending on whether one is working within a primarily representational or primarily derivational model, one finds different kinds of economy considerations in the literature. The overview here will focus on three main kinds of economy: global derivational economy, local derivational economy, and global representational economy. Then, I will focus specifically on economy in the Minimalist Program, moving to a detailed consideration of one economy principle of the MP in particular: Merge over Move. I would argue that this is, in fact, the principle most relevant to economy in the MP, and that furthermore it is the most suspect. Section 2 is a meta-linguistic critique of the variety of instantiations of ‘economy’ in the MP; Section 3 narrows the focus to Merge over Move, and presents critique and discussion of the core empirical case that has been made in favour of this principle. This section presents and evaluates the alternative ways in which this core empirical argument can be accounted for without the economy principle Merge over Move; there is

---

1 I leave aside the question of whether ungrammatical forms can be said to be strictly “ruled out,” or whether they are simply not generated.

2 When I use “derivations” here, I do not mean to exclude representational systems.
discussion of the empirical arguments against the principle as well. From there, Section 4 takes up residual, more minor empirical arguments that have been made for the Merge over Move principle, and deflates each of these with suggestions for alternative analyses. The final section discusses the theoretical consequences of eschewing the economy principle MOM, and concludes by relating back to the general issue of economy principles in the MP.

1.2 Economy in different frameworks

Economy has been formulated in many different ways in many different frameworks. We can identify several major types of economy: global derivational economy, local derivational economy, and global representational economy.

Global derivational economy looks at an entire derivation as a whole and considers whether the resources involved in its computation are greater or less than the resources involved in another derivation that begins with the same numeration and converges. Global economy thus often comes down to counting steps of derivations: the derivation constituting the ‘shortest’ path from the lexicon to the interface levels is ‘preferred’ by the grammar in some way, to the point that the less economical of two derivations can actually be excluded for this reason alone. Take, for example, the strong global economy principle from Kitahara (1995): the Shortest Derivation Requirement.

\[\text{Minimize the number of operations necessary for convergence.}\]

Steps are usually defined by the number of operations performed, but it is not always uncontroversial to decide which operations should be counted (or what ‘counts’ as an operation). Global derivational economy also entails that the grammar has a ‘lookahead’ ability; in other words, it entails that at a particular point during a given derivation, the grammar can look ahead to see if the derivation will converge. Sometimes theories of global economy also allow the grammar to compare one derivation to other possible derivations, in order to pick the least costly one. For global economy that counts steps, the grammar also has to be able to compare properties of derivations (i.e. their length). The lookahead problem has been a fairly large concern of theorists, since it greatly increases the computational complexity of derivations; however, Juan Uriagereka (p.c. March 2009) notes that there is a sort of compromise version between global and local economy (which does not espouse lookahead)—a sort of “local-global economy”—which might not be so taxing on the computational system of the grammar. I will not be concerned much in this paper with issues of computational complexity; those issues have been addressed elsewhere and by those who have more expertise in discussing such issues. I will be more concerned here with the empirical basis for economy principles, and the theoretical motivation behind them.

Representational economy begins with different assumptions about narrow syntax than derivational economy does, endorsing the idea that it is some kind of economy of representations, rather than derivations, that is active in the grammar. One example from the literature of this kind of economy is Williams’s (2003) Representation Theory, where Williams argues that the grammar works to economize based on ‘shape conservation.’ Briefly stated, shape conservation takes the idea that there are different levels of representation, such as Case structure, Thematic structure, Quantifier structure, etc., and that where mismatches between these levels occur, the grammar chooses the structure that causes the least distortion to each of the various levels of representation. The terms in which this theory is couched are strictly representational. I will not be working within such a strictly representational system in this paper, instead focusing on the ‘mixed’ system (involving derivations and representations, although primarily the former) that is instantiated in the Minimalist Program. Though this paper will have comments relevant to both global derivational economy and local
derivational economy (discussed below), representational economy starts with such different assumptions about the basic structure of the syntactic computational system that a discussion which does it justice is not within the scope of this paper.

Lastly, local derivational economy works within a derivational system, but unlike global economy, it does not (typically) involve comparing alternative derivational paths, nor does it generally entail that the grammar has a lookahead property. For this reason, it is often preferred on the theoretical grounds that this greatly reduces the complexity of the computations that the grammar must perform. Collins (1997) has argued for a strictly local economy with substantial empirical coverage. Rather than comparing convergent derivations, local derivational economy principles compare options at a particular point of a derivation and decide which derivational path is most economical. Local economy principles are generally blind to further consequences of choosing one path or another; for example, such principles might not know whether a particular path converges or not. These principles must therefore make decisions about derivational economy based only on the information available at the point in the derivation where the economy principle applies (or at least this is true for the strictest interpretation of local economy). Sometimes it may not be explicit whether an economy principle is global or local, or a principle may live a ‘double life’ and be interpreted as both global and local. Some instantiations of the Merge over Move principle, discussed at length in Sections 3 and 4 of this paper, are like this.

In this paper, I will primarily deal with ideas of economy within the Minimalist Program, not because economy in the other frameworks does not merit equal scrutiny, but for two reasons: first, a close examination of economy in the MP is more than enough to cover for a project of this size, and second, this is the framework in which I think the most serious problems of consistency in defining economy are found.

2. Economy à la carte in the Minimalist Program

2.1 Substantive versus methodological economy

Let us make a distinction, first, between methodological economy and what Hornstein et al (2005) call substantive economy. Methodological economy is what guides us in postulating the simplest possible explanation of a phenomenon. By extension, this means that we ought to have the simplest scientific model that adequately captures all the relevant facts (leaving aside the problematic issue of what ‘all the relevant facts’ are). This is the idea popularly known as Occam’s razor: the idea that superfluous postulates are dispreferred, assuming all else is equal. This general preference towards simplicity and parsimony is appropriately applied to models of phenomena in science, and inappropriately applied to the systems themselves that we are attempting to model—unless, of course, one has evidence that some kind of economy is active in the system itself. When we apply this general discussion to Universal Grammar (UG), then this second kind of preference for parsimony, not in the model but in the thing being modeled, the kind that is motivated by empirical data, is what we will call substantive economy. Hornstein et al (2005: 8) describe this second kind of economy as the practice of “placing a premium on least-effort notions as natural sources of grammatical principles.” The hypothesis that UG itself is based on principles that favour more economical operations, derivations, etc. derives from Chomsky (1991: 130):

“I think we can also perceive at least the outlines of certain still more general principles, which we might think of as ‘guidelines,’ in the sense that they are too vaguely formulated to merit the term ‘principles of UG.’ Some of these guidelines
have a kind of ‘least effort’ flavour to them, in the sense that they legislate against ‘superfluous elements’ in representations and derivations.”

Such considerations in the grammar, if they are present at all, are outside the range of methodological economy. The kind of ‘guidelines’ Chomsky refers to in the above passage are principles within the grammar itself, not guidelines for constructing a parsimonious model of the grammar. However, in the above paragraph, the line between methodological economy and derivational economy is somewhat blurred. When Chomsky states that these least-effort guidelines “legislate against ‘superfluous elements’ in representations and derivations,” he is referring to the grammar, but using terminology that is equally applicable to the model. This will be discussed again in a later section.

Some might argue that this distinction between the grammar and the model of the grammar is trivial, or that such a distinction is unnecessary since the model exactly replicates the grammar. (Or, perhaps, the grammar can be thought of as no more than what we know it to be—in other words, our model.) I disagree: the distinction is not trivial, and methodological economy should concern us only insofar as we are dealing with the methodology of constructing our models. Moreover, concerns about parsimony in UG which have the flavour of these methodological concerns (that is, they are substantive concerns posing as methodological concerns) should not fool us. If the motivation for substantive economy is primarily empirical, then the proof is in the data. We do not need to, and should not, look to the Occam’s razor concept of methodological economy to justify any principle falling under the banner of substantive economy. This is a misapplication of the concept of methodological economy. Conversely, concerns about superfluous elements in our representations or our derivations most correctly fall under the banner of methodological economy, and should not concern us when we are trying to pinpoint principles of substantive economy, unless we have other reasons to assume such economy within the system (UG) itself.

On the other hand, if substantive economy is real, then its nature is a matter for empirical investigation. It doesn’t have to toe the party line of “legislating against superfluous representations and derivations”: if the grammar is economical, the metric of economy, and the range of economy will be defined only by the empirical evidence. In other words, economy principles that are substantive should be motivated substantively. Minimality (or the Minimal Link Condition), or “Shortest Move,” is an economy principle I would hold up as a paragon of a well-supported substantive economy principle. Though the various specific instantiations of the idea may differ from theory to theory, or from author to author, the general idea that a derivation selects and moves components that are closer to the target rather than those that are farther, has proven resilient. But not all economy principles are equally robust, and the fact that they are called ‘economy principles’ furthermore should not influence our responsibility to demand substantial empirical verification of them.

In this way, it is possible to imagine how methodological economy and substantive economy could be in a kind of tension regarding one another. Substantive economy principles might be introduced based on data unexplainable by other means; methodological economy concerns, on the other hand, naturally ‘legislate against’ the introduction of new principles, unless those principles are proven necessary to capture all the relevant facts. One of these kinds of economy (substantive) tends towards the licensing of principles that can be defined by some kind of economic metric; the other kind of economy (methodological) naturally tends towards ruling out unnecessary postulates, including principles and other stipulations which do not do any necessary work in the theory, or do work that could just as well be done by already existing principles. (Which is not to say that already-existing principles are privileged in any way over innovative ones—merely that a smaller set of principles is naturally preferred by this second kind of economy.)
Furthermore, the kinds of economy principles that have been proposed display a kind of patchwork, ad hoc quality, which is the motivation for the title of this section. I will argue that this quality of economy principles, taken as a whole, is symptomatic of a lack of empirical evidence for certain kinds of substantive economy as an active mechanism in UG. It is necessary at this point to carefully define my position on economy. First, assuming economy is not a functional mechanism within the grammar does not prevent it from being predictive (and therefore useful). Even if economy is epiphenomenal to the grammar, certain economy principles may have striking predictive power. (It is in fact my intuition that several economy principles that have been proposed are reflections—that is, are derivative—of other active processes or combinations of processes in such a way as to appear to act as principles of the grammar when they are in fact only observationally adequate generalizations.) So this is not an argument against the reality of economy as observationally relevant to the way the language faculty is organized. It is also not an argument that there are no economy principles active in the grammar. It seems obvious to me that something like Relativized Minimality (or a locality principle of some kind) is very well supported empirically. Other ‘economy’ principles, however, are not. Merge over Move will be the case study for one such principle in this paper. So instead of an anti-economy argument, this is actually an argument (or the beginnings of an argument) against only certain substantive economy principles as functional mechanisms within the grammar. It is also an argument for careful empirical support in the domain of economy principles, and somewhat of a warning against defining economy over too many disparate metrics just because we can. If we want economy to be a functional mechanism in our grammar, we must show what it does; furthermore, if we want to be methodologically economical (which we do), then we shouldn’t introduce any principle, whether it be an economy principle or not, that we don’t need.

There are other kinds of arguments made in the domain of substantive economy that do not rely on empirical motivation: there are, for example, more theory-internal arguments for specific principles of substantive economy, some of which will be dealt with later in this paper. I will put off further discussion of such arguments to later in this paper, when we are more equipped to hold such a discussion. What follows is a brief introduction to the origins of economy in the MP.

2.2 The origins of economy in the Minimalist Program

2.2.1 Chomsky and Lasnik (1993)

The idea of economy is first made explicit in Chomsky (1991), “Some Notes on Economy of Derivation and Representation.” However, reference to notions of economy also occurs in other sections of The Minimalist Program (Chomsky 1995), for example, in, “The Theory of Principles and Parameters,” written with Howard Lasnik (published as Chapter 1 of The Minimalist Program, originally published in Jacobs et al (eds.) 1993, Syntax: An International Handbook of Contemporary Research. This chapter is referred to here as Chomsky and Lasnik 1993). Chomsky and Lasnik (1993: 23) say the following:

“The principles [of UG] have further structure […] There are also certain general ideas that appear to have wide applicability, among them, principles of economy stating that there can be no superfluous symbols in representations (the principle of Full Interpretation, FI) or superfluous steps in derivations.”

---

3 The title I owe, in fact, to Alexander Williams, p.c., in a very brief conversation that I would not be surprised if he had already forgotten!
It is evident that the exact metric of economy has yet to be defined at this point, since the above quotation is followed by the remark, “As these principles are given an explicit formulation, they become empirical hypotheses with specific import and range.” Crucially, economy here does not have “specific import and range” and is not an empirical hypothesis. It will be part of the project of this paper to identify how far we have come from this statement, and how successful we have been at identifying and validating empirical hypotheses of economy. Chomsky and Lasnik (1993: 28) provide a vague indication of what they imagine a fully understood concept of economy in the grammar might look like: “Interacting with other principles of UG, […] economy principles have wide-ranging effects and may, when matters are properly understood, subsume much of what appears to be the specific character of particular principles.” It thus appears from this that the expectation for economy, at this point in the theory, is for it to act as a large, unifying principle that potentially lies behind (or above) other more specific principles.

More specific examples of economy are invoked elsewhere in Chomsky and Lasnik (1993: 69):

“Suppose further that economy principles favour operations that do not feed the PF component over others that do; hence, if operations need not be overt to satisfy some condition, they will be assigned to the LF component, applying as “late” in the derivation as possible, at the point where they are forced by LF conditions […]”

The empirical importance of such a ‘flavour’ of economy is that in conjunction with the postulated Q (question) feature, it allows for a systematic explanation of why some languages necessarily raise wh-words (question words) in question formation, before Spell-Out, while others do not. The background to this is the oft-observed distinction between languages that leave their question words in situ, like Chinese, and those that move one of their question words to a high scope position (standardsly [Spec CP]) before Spell-out (English-type languages), thus overtly. In Chinese-type languages, all question words move at LF, but covertly, after the divergence to PF of that portion of the derivation. In English-type languages, any additional question words (for example in sentences with multiple question words—e.g. What did John buy for whom?) move only after Spell-out, thus covertly. There is also a third type of language (exemplified by Polish) that moves all question words overtly, or before Spell-out. The distinction as described above rests on the assumption that all languages move their question words to a position of high scope at least by LF; the question is then why some languages move these words before LF. This leads to the postulating of a feature Q. In Chomsky and Lasnik (1993) it is the absence of Q in a language’s lexicon that causes movement to occur pre-Spellout, though this is revised in later theory, which understands there to be both weak Q and strong Q features (strong features motivating movement to occur pre-Spellout). Why then do we need to invoke some kind of economy as well?

Without an economy principle to delay unnecessary movement to the ‘moment’ (within the derivation) when it is necessary, the theoretical concepts thus far supplied are insufficient to derive the attested patterns. This is the state of the explanation without economy:

(2) a. All Q-phrases must move to high scope positions before LF.
    b. If a Q feature is present, Q-phrases don’t have to move before Spell-out.
    c. If a Q feature is not present, one Q-phrase must move to [Spec CP] before Spellout to meet the Condition on Mood at S-Structure.\(^4\)

\(^4\) As in Cheng (1991): the idea that the mood (e.g. interrogative) “must be indicated at S-structure in the pre-IP position, hence by a choice of either C or [Spec CP]; the head of CP and its specifier thus serve as “force indicators” in something like the Fregelian sense,” (as cited in Chomsky and Lasnik 1993: 69).
The problem is with assumption (2b): there is no condition stating that without a Q feature, not only do Q-phrases not have to move before Spell-out, but they won’t. Without this stipulation, we would expect that the presence of a Q-feature would not necessarily mean that Q-phrases wouldn’t move before Spell-out: it could then be that a Q-feature is present, and the question word still moves higher in the tree pre-Spellout. This loses the one-to-one relation between a Q-feature being present and some definable effect following on the location of question-words in a sentence. Some kind of principle is thus needed here to ensure that if question-words aren’t required to move, they won’t, at least until they do have to, and this is where economy, or some manifestation of it, steps up to bat.

Now it is interesting to consider whether it must be economy that saves explanatory adequacy here. Certainly, it would be enough to stipulate that the presence of the Q-feature prevents the Q-phrase from moving before Spell-out. Is this really ‘economy’ or is this just part of the mechanism of how features work? For Chomsky and Lasnik, as indicated in the quotation from page 69 given above, this is an instantiation of a larger theme in the syntax: an operation will be delayed until as late as it can, or until the moment when it is required. This is an early illustration of the principle Procrastinate. Procrastinate essentially makes LF movement (covert movement) preferable to overt movement; however, within the current framework, Procrastinate is reducible to Last Resort (Collins 2001), another economy condition which, simply put, states the following:

(3) **Last Resort:** An operation OP may apply only if the derivation would otherwise result in an ungrammatical representation (at PF or LF). (As stated in Collins 2001: 46)

If we infer from the above definition that an overt operation is only admissible if the derivation would otherwise crash, it follows that covert movement, which does not involve overt operations, will be less costly in the grammar. But how is this economy?

More specifically, what makes it like the ‘economy’ of the Minimal Link Condition, for example? Is there something that these share, beyond our speculative intuitions that doing something ‘late’ in a derivation conserves the resources of the computational system of the language faculty in the same way that selecting a goal that is ‘closer’ in the syntactic tree under construction does? If we believe these are both manifestations of the ‘economy’ of the grammar, it also seems we are assuming quite a lot about how the grammar uses its resources, and what cost is apportioned to such use. There is no imperative to make the leap from the fact that features will be checked covertly if possible, overtly if necessary, to the idea that this is unified, by economy, with the principle of Minimality, to take just one example.

Economy also arises later in one other case in Chomsky and Lasnik (1993: 89 - 90), in the context of the locality of government. Economy is invoked here, again, as a kind of unifying principle that subsumes smaller, more detailed principles. The idea here is to reduce the complexity of the formulation of the notion of government. A tentative definition of government is laid out in Chomsky and Lasnik (1993: 79): α governs β if α c-commands β and there is no intervening category γ such that γ is c-commanded by α, and either of the following holds: a) γ is a barrier dominating β; b) γ intervenes between α and β. The nature of γ, and whether it can be an intervening category or not, depends on the Minimality Condition. After considering briefly Rigid versus Relativized Minimality, Rigid Minimality is rejected. Relativized Minimality specifies that γ, to be an intervener, is of the same ‘type’ as α, with the relevant types being three-fold: i.) α and γ are heads, or ii.) α is in an A-position, and γ is a specifier in an A-position, or iii.) α is in an A’-position, and γ is a specifier in an A’-position. However, Chomsky and Lasnik’s complaint about this formulation of government is that while it obtains descriptive adequacy, it fails to achieve explanatory adequacy: “These observations have a wide range of descriptive adequacy, but fall short of a satisfactory explanatory principle” (1993: 82). The problem is that, according to the authors, the
mechanisms proposed above fail to be methodologically economical: “the basic and appealing intuition that lies behind the principle of Relativized Minimality is not really captured by the mechanisms proposed, which list three arbitrary cases and add unexplained complexity” (1993: 89).

The authors thus propose to simplify the formulation of government by discarding these three classes of interveners, since movement of XPs to specifiers and heads to head positions is restricted independently of the formulation of government. They are thus left with the condition: Minimize chain links. This is a preferable principle to postulate in the grammar because it is much simpler. I would suggest that this is an appropriate way to appeal to methodological economy. The authors then generalize to hypothesize that, “The basic intuition is that the operation Move α should always try to construct ‘the shortest link.’ If some legitimate target of movement is already occupied, the cost is deviance […] We may regard this as part of the general principle of economy of derivation” (89-90, italics mine). Here, the move has been made from reducing the complexity of the principles of the grammar by subsuming smaller principles under a broader one (a valid instance of methodological economy considerations), to the speculation that this is a part of some larger, umbrella-like concept of economy of derivation. Crucially, not only is this larger, broader appeal to a general notion of economy of derivation a vague hypothesis, but it is a hypothesis about a principle of the grammar. I assume the idea of a principle here implies a condition that has an active function in the grammar, and indeed that appears to be the case in Chomsky and Lasnik’s hypothesis about the principle, Minimize chain links (1993:90):

“To make this intuitive account more precise and descriptively more accurate, we have to explain in what sense a ‘cost’ accrues to failure to make the shortest move […] We might suppose that when a chain link is formed by Move α, the trace created is assigned * if the economy condition [Minimize chain links] is violated as it is created.”

“Minimize chain links” is an active principle since its violation has an effect on the grammaticality or ungrammaticality of a form (thus whether a form is generated). It makes empirical predictions. Note also, though, that the language in which the authors appeal to the more encompassing notion of derivational economy is also professed to be a principle, thus active in the grammar, and making empirical predictions. While “Minimize chain links” is testable empirically, it is doubtful that “the general principle of economy of derivation” is clearly defined enough to be equally as easy to empirically validate or invalidate. This is at least in part because it is not fully defined (the more specific economy principle, Minimize chain links, only defines ‘a part of’ it), and until it is, the “general principle of economy of derivation” is more of a unifying idea than a unifying principle (the distinction being that the latter does identifiable work in the grammar).

2.2.2 Chomsky (1991)


“I think we can also perceive at least the outlines of certain still more general principles, which we might think of as ‘guidelines,’ in the sense that they are too vaguely formulated to merit the term ‘principles of UG.’ Some of these guidelines
have a kind of ‘least effort’ flavour to them, in the sense that they legislate against ‘superfluous elements’ in representations and derivations.”

The two particular principles Chomsky refers to at this point are Full Interpretation and Last Resort, with the intention that the former restrict any superfluous constructs in representations, and the latter have “the corresponding effect of eliminating superfluous steps in derivations, thus minimizing their length” (130). Neither of these is exactly a principle, and this deficiency is acknowledged in the next lines:

“What I would like to do here is to search for some areas where we might be able to tease out empirical effects of such guidelines, with a view towards elevating them to actual principles of language, if that is indeed what they are.”

The idea of a ‘general principle of economy’ is still elusive, and the methodology at this point is to assume such guidelines exist, though their exact form is still vague, and search for places in the grammar where they can be supported by empirical facts. It is possible that this general methodology biases the researcher toward an overrepresentation of economy in the grammar—a sort of ‘wherever we find it, we’ll take it’ approach. The danger in starting with an ‘umbrella’ idea of economy and looking for ways to substantiate it (as Hornstein et al 2005 state, the goal of, “putting linguistic flesh on methodological bones”) is that one might find ample opportunity of suiting the economic metric (or the specific definition of an economy principle) to whatever best suits the current empirical situation. In the above quotation, I also feel there is a blurring of the line between methodological and substantive economy. It is possible that Occam’s ideal has been generalized to the grammar itself in a way that is more convenient than it is true: while we look for the empirical facts to tell us what substantive economy principles hold in the grammar, we assume a kind of nebulous, unifying economy ideal which borrows in great part from our methodological concerns as theorists.

Two kinds of economy are explicitly discussed with empirical evidence in Chomsky (1991): shorter derivations, with the French verb-raising paradigm as an example; and the idea that UG rules are less costly than language-specific rules, with English do-support providing the empirical support.

2.3 ‘Flavours’ of economy: any way we want it!

A crucial part of my criticism of economy in MP is that once we assume economy is something that we like, something that we want to be true of the grammar, and once we accept the strong hypothesis that the grammar itself is optimal, we go too far with this idea and start defining economy in all sorts of ways. Furthermore, the kind of economic metric used in individual cases, whether it be distance or operation strength, is often defined post hoc—that is, after we’ve seen what kind of economic metric could be used to solve the particular problem at hand. In this way, economy often arises as an ad hoc solution to very specific phenomena.

Below, I outline a very general typology of the uses of economy within the MP, and instantiate these with examples. In Section 3 I will then go into more detail on one instantiation of the last kind of economy in the typology, economy of operation strength, in the Merge over Move principle. Section 3 is a detailed analysis of the disputed empirical necessity of this principle. Note that Collins (2001) has a similar outline of the various types of economy.

5 Note that Chomsky (1991) refers the reader to Chomsky (1986) and Chomsky (1951) for further, earlier discussion of these issues.
2.3.1 Economy as distance

‘Least effort’ can mean that within a single representation, at a single point in the derivation of a particular structure, a goal that is closer to a probe will block a goal that is farther away. A very simple example of this kind of economy in action is the following:

\[
\begin{align*}
(4) & \quad \text{a. } [\text{could}, [\text{they t}, [\text{have left}]]] \\
    & \quad \text{b. } [\text{have}, [\text{they t}, \text{left}]] \\
    & \quad \text{c. } [*\text{have}, [\text{they could t}, \text{left}]] \\
\end{align*}
\]

(data from Hornstein et al 2005: 144)

In (4) above, \textit{have} is only allowed to move into the CP domain if there is no other auxiliary head that is closer that could do the same (i.e. \textit{could} in [4a]). The idea of ‘shortest move,’ in some form or another, is very prevalent in earlier theories of grammar (i.e. before Minimalism) as well as recently, and it seems to have substantial empirical support in the literature. Note that Minimality like that in (4) above is category-sensitive; that is, a head can block a head if it is closer to the probe, and a DP can block a DP if it is closer to the probe, but a DP cannot block a head. This kind of economy is quite representational. The principle measures distance in terms of the amount of structure (i.e. number of projections) between probe and goal. The idea of ‘shortest’ here is to a certain extent metaphorical: we know little about how the grammar ‘reads’ that intervening structure.

2.3.2 Economy as feature strength

The idea of strong and weak features has also been expressed as a kind of economy. Features trigger movement: weak features trigger covert movement, and strong features trigger overt movement. This is the same overt/covert distinction that the Procrastinate principle is meant to capture. However, one might as well ask whether this really captures something economical in the grammar, or whether it’s just an idiosyncratic quality of the mechanism by which feature checking occurs. Certainly, the notational labels ‘weak’ and ‘strong’ imply that one causes the grammar to expend more resources than the other, but the essential idea of some features being checked before Spell-out and some after can be expressed in terms that are economy-neutral.

2.3.3 Economy as derivation length

The length of a derivation is generally measured by the number of ‘steps’ in the derivation; of course, the number of steps we count relies entirely on what we define as a step. Though we can count steps, it is not entirely obvious that the grammar does, though this has been argued. In general, one step is assumed to be equivalent to one application of a core operation, (e.g., Merge or Move), but there is also controversy about what constitutes a ‘core’ operation. This will be discussed further below and, for Merge, Move, and Copy, in great detail later in this paper. The idea that the grammar counts steps of a derivation has not had much support in recent years, perhaps for the computational burden it would seemingly add to the grammar. This kind of economy generally works in the following way: shorter derivations block longer ones. The longer derivations (i.e. with more steps) will be ungrammatical, though in the grammar they may actually be (theoretically) convergent.

2.3.4 Economy as effort

Certain economy principles seem to attribute computational burden to a derivational path based on complexity of the components being copied or moved. For instance, the ‘Weight Condition’ mentioned in Collins (2001: 56) states that “F carries along just enough material for convergence.”
According to Collins, Chomsky proposes this principle in order to “regulate […] generalized pied-piping.” This is similar, if not identical, to the “Lightest Load principle,” which Hornstein (2009: 40) attributes to Chomsky (1995).

2.3.5 Economy as timing

There are economy principles that propose that certain timing preferences of the grammar are economy-based. Procrastinate and Last Resort delay operations from occurring until they must. An “As Soon As Possible” Principle (ASAP) has also been proposed (see Yang 1997, Collins 1999, Collins 2001):

(5)  As Soon As Possible Principle: If it is possible for an operation to apply, then it must apply.

It is assumed that these two kinds of principles, those that catalyze operations and those that delay them, are not in inherent contradiction with each other since they both simplify the operations that the computational system of the grammar needs to perform.

2.3.6 Economy as operation complexity

Chomsky (1995, inter alia) has argued that certain core operations of the grammar are available ‘for free,’ while others ‘cost’ the grammar an amount of its finite resources. Merge is argued to be free because any theory of grammar requires, by “virtual conceptual necessity”, an operation that concatenates elements. Move is generally argued to be more complex, and therefore more costly. These particular operations will be extensively discussed in the rest of this paper.

2.4 Summing Up

As I see it, the problem exemplified above is that economy has arisen in too many disparate ways in the theory of universal grammar. I think this has primarily been the result of a top-down analysis that starts with the assumption that some kind of economy conditions are present in the grammar, and then goes on to look for the right way to define that economy—the right ‘economic metric’—in order to make particular empirical facts in a certain case study fall out correctly. I think it should be fairly apparent, conceptually, that not all of the above ‘flavours’ of economy can be unified under one banner. They are different concepts requiring different kinds of empirical and conceptual motivation. So it is not facetiously that I ask the question, what is your favourite flavour of economy?

This is not to say that economy does not exist in the grammar at all; however, I hope that the discussion in this paper will motivate a degree of healthy skepticism in formulating ‘economy conditions’, and a further fortification of the need for strong empirical support for such conditions as they are proposed. A further warning might be to be wary of what we call ‘economy.’ It might be that several sound principles of the grammar, each with its own portfolio of empirical support, can only be unified under the banner of ‘economy’ insofar as our imaginations succeed in stretching this metaphor along the convenient metric.

The rest of this paper will primarily address one particular economy principle, the empirical evidence for which is particularly scarce. Merge over Move was first proposed by Chomsky (1995: Ch. 3) based on one specific kind of empirical data from English, discussed at great length in the section below. The principle has since become suspect, since nearly a decade and a half’s subsequent investigation by researchers in a great number of languages has failed to uncover any substantial new evidence for the principle. In fact, subsequent research has rather tended to find evidence against the principle, or to find that it is unnecessary to account for the core empirical case. Despite this, as we
shall see in the final section of this paper, much theoretical apparatus has subsequently been built upon the assumption that Merge over Move plays an active role in the grammar. Though a thorough investigation of the theoretical implications at stake in this principle is not the primary aim of this paper, the possible ramifications of rejecting the principle are discussed in Section 4. Section 3 details the core empirical case for Merge over Move from Chomsky (1995), and presents several key responses to it from the recent literature. This section also reviews the alternative analyses for the core empirical case which have been proposed, with the aim of giving the reader enough information to decide for him/herself whether the Merge over Move principle is really necessary to account for this core empirical case. My own opinion is that it is not. Merge over Move is in a uniquely precarious situation: while it appears to lack very substantial empirical evidence, it has nevertheless been a cornerstone to much recent theoretical apparatus, such as subarrays and phases. Merge over Move has also received a great deal of—mostly critical—attention in recent years, one of the reasons this paper chooses to focus on it. As far as economy principles go, the evidence for this one is weak.

3. Merge over Move: An economy principle up close

3.1 The original empirical case for Merge over Move

The most significant debate over an economy principle has been waged over the Merge over Move principle. Simply put, Merge over Move (MOM) stipulates that between two comparable derivations, the grammar favours the merge of a new element from the array at a given point in both derivations, rather than moving an element already in the structure being built. The grammar in fact favours this first derivation to such a degree that the derivation which chooses to move an element already in the structure, rather than selecting a new element from the numeration, will be ruled out by the principle, since a more ‘economical’ derivation exists. Why is the “merge” derivation more economical? Arguably, it’s anyone’s guess, but what has been proposed in the literature is that the operation Merge is more economical than the operation Move because Merge is inherently less complex than Move. This story rests on some very specific stipulations about what makes two derivations comparable. The idea is explained in detail below. Merge over Move is thus a specific instantiation of the “general principle of economy of derivation” that has been at issue in earlier sections of this paper.

The crucial data are existential sentences of the following kind:

(6) a. There seems to be someone in the room/here.
b. *There seems someone to be in the room/here.  

The contrast between (6a) and (6b) is problematic in the Minimalist framework. The derivation proceeds as follows, deriving sentence (6b) apparently without any difficulty. From the numeration in (7a), the operation Select selects the most deeply embedded elements from the numeration, someone and here, and the operation Merge concatenates them to arrive at (7b). The numeration is then in state N1, indicated in (7c). The operations Select and Merge then operate again on the remaining elements of the numeration, and (8a) is formed, with the resulting numeration being (8b). Another round of select and merge derives (9a) and the corresponding numeration (9b). Further steps in the derivation of (6b) are delineated in (10) – (13) below.

---

6 Note that the same contrast occurs with any expletive sentence that embeds an infinitival clause under a raising verb. For example:

(i.) a. There is likely to be a rainbow after the rain.
b. *There is likely a rainbow to be after the rain.
So, we see the ungrammatical sentence, *There seems someone to be here*, is derived without any difficulty. This sentence is quite robustly ungrammatical, so there must be something missing from our explanation of how this sentence is derived. The solution suggested by Chomsky (1993), “A Minimalist Program for Linguistic Theory” (subsequently published as Chapter 3 of *The Minimalist Program*) is that there is an additional economy principle, Merge over Move, which is factored into the derivation of this sentence. Note that, as with the discussion of question-word raising from Chomsky and Lasnik (1993), the ‘discovery’ of this need for an economy principle follows from the inability of the current theoretical apparatus to account for particular empirical facts. The economy principle is added to account for the facts, which naturally puts it into tension with methodological economy. Of course, this is exactly how we want to have substantive economy principles be motivated: introduced where necessary to account for empirical patterns of evidence. However, the tension between methodological and substantive economy exists even when economy principles are empirically motivated. The question remains whether it is possible to account for the aberrant pieces of data without adding an economy principle. This question is particularly salient when the empirical evidence for the principle is not overwhelming (as, I will argue, is the case for Merge over Move).

Economy is introduced into the picture in the following way: there is a second way to get a convergent derivation using the same elements specified in the numeration $N_0$. This derivation is identical to the previous derivation for *There seems someone to be here* up until the stage (9a), repeated below as (14):

(14)  

a. $[\text{TP to be someone here}]$  
b. $N_3 = \{\text{there}_1, \text{T}_1, \text{seem}_1, \text{to}_0, \text{be}_0, \text{someone}_0, \text{here}_0\}$  

---

7 Note that I am not committed to the label of this phrase (SC = small clause). It does not particularly matter to this analysis what kind of phrase (2b) is.
At this point in the derivation, there are two options available to the syntax. One of these options was outlined in (10) – (13) above: someone moves up from lower in the embedded clause. The second is to take there from the numeration and merge it into the structure in [Spec TP]. Both of these options satisfy the derivation’s need to check the EPP (Extended Projection Principle) feature (or strong D feature) of T0 at this point in the derivation. However, only one option will lead to a de facto grammatical output—the option that chooses to merge there at this point rather than waiting until later in the derivation to merge it. If at this point we merge there into the structure, the derivation proceeds as follows:

(15)
\[
\begin{align*}
a. & \quad [\text{TP} \text{there to be someone here}] \quad \text{(MERGE)} \\
   b. & \quad N_4 = \{\text{there}_0, T_1, \text{seem}_1, \text{to}_0, \text{be}_0, \text{someone}_0, \text{here}_0\}
\end{align*}
\]

(16)
\[
\begin{align*}
a. & \quad [\text{VP} \text{seem} [\text{TP} \text{there to be someone here}]] \quad \text{(MERGE)} \\
   b. & \quad N_5 = \{\text{there}_0, T_1, \text{seem}_0, \text{to}_0, \text{be}_0, \text{someone}_0, \text{here}_0\}
\end{align*}
\]

(17)
\[
\begin{align*}
a. & \quad [\text{TP} T [\text{VP} \text{seem} [\text{TP} \text{there to be someone here}]]] \quad \text{(MERGE)} \\
   b. & \quad N_6 = \{\text{there}_0, T_0, \text{seem}_0, \text{to}_0, \text{be}_0, \text{someone}_0, \text{here}_0\}
\end{align*}
\]

(18)
\[
\begin{align*}
a. & \quad [\text{TP} \text{there} T [\text{VP} \text{seem} [\text{TP} \text{there to be someone here}]]] \quad \text{(MOVE)} \\
   b. & \quad N_7 = \{\text{there}_0, T_0, \text{seem}_0, \text{to}_0, \text{be}_0, \text{someone}_0, \text{here}_0\}
\end{align*}
\]

As we can see from the output of this alternative derivation (6a), *There seems to be someone here*, merging there earlier in the derivation (in the embedded clause) results in a grammatical output. Both of the derivations we have examined here are convergent (they ‘work out’), so both should result in grammatical sentences of English. This is not the case, though, as we observe from the contrast between (6a) and (6b). Moreover, both derivations have the same number of steps, which satisfies global economy. But even if we didn’t believe in the kind of global economy that keeps track of how many steps are in each of two comparable derivations, the two derivations depart at the same point (as noted by Hornstein et al 2005: 337), so local economy will serve our purposes here. In any case, we can’t differentiate the two derivations based on counting steps of the derivations, so another metric of economy must be defined if we want to differentiate these derivations based on some notion of economy. Thus we come to the notion of economy of operation complexity. This is an entirely new notion of economy, a different ‘economic metric,’ as one might say, and it allows us to explain why *There seems someone to be here* is ungrammatical, while *There seems to be someone here* is grammatical. The former is ungrammatical because the economy principle Merge over Move is consulted and the two possible convergent derivations compared: Merge over Move then “rules out” the derivation which Moves instead of Merges at the comparable point, thus the sentence using this dispreferred operation is ungrammatical.8

Returning to the issue of what makes these two derivations comparable for purposes of economy: for one alternative derivation to block another based on an economy principle, the two alternative derivations must start from the same numeration, and both must be convergent. This is the case under the above analysis of the data in (6): both start from numeration N0 (above), and both converge given the derivations (and assumptions) outlined above.

---

8 The comparison of Merge and Move based on operation complexity is local: both the derivation of (6a) and the derivation of (6b) have the same number of Merge operations and the same number of Move operations in total (as indicated in [10] – [13] and [15] – [18]). However, though the decision of which operation to apply is made locally (i.e. at the step in [9][14]), Merge over Move will only choose between convergent derivational paths.
3.2 The theory-internal case for MOM: the proper-subpart argument

There is a theory-internal reason for believing this explanation, in addition to the empirical story: Merge is supposedly inherently less complex than Move. I will call this the “Proper Sub-part Argument,” based on terminology used in Hornstein et al (2005). The argument goes like this: since Merge is a “proper subpart” of Move, the operation Merge is inherently less complex than Move. Move, under this analysis, is a composite of two more atomic operations: Copy and Merge. If Move contains Merge as one of its sub-components, then Move must be a more complex operation. This argument for Merge over Move rests on the assumption that operational complexity in this sense translates directly into ‘more cost to the grammar,’ so that, in the flavour of economy mentioned earlier—operational complexity—it is more economical to choose, at a local point, a less complex operation over a more complex one. As Chomsky (1998: 138) states:

“It was argued [in the MP] that Merge is preferred over Move, but on dubious grounds. Now the preference is immediate: Agree and Merge are each components of Move, so it is a simple matter of more versus less.”

Shima (2000) recaps this motivation in detail:

“Move (a) establishes agreement between a lexical item α and a feature, F, (b) selects P(F), and then (c) merges P(F) to αP, where P(F) is a phrase determined by F and αP is a projection headed by α. Since Move involves the extra step of selecting P(F), Move is more complex than its subcomponents Merge and Agree, or even the combination of the two. Given that simple operations preempt complicated ones, Merge is preferred to Move.”

In other words, Chomsky (1998) argues that both Agree and Merge are simpler than Move, though based on slightly different sub-components. The way this is presented in Hornstein et al (2005) involves Copy as one of the subcomponents of Merge, but Chomsky (1998) uses slightly different steps. Step (a) above of Move is essentially equivalent to the operation Agree; step (c) is essentially the operation Merge; Move is the combination of these first two plus the additional step (b), and thus is inherently a more complex operation. Even the combination of Merge and Agree would yield a composite operation of only two of the above-identified three steps, and so such a composite operation would still be less complex than the tripartite operation, Move. A critical examination of this argument involves an evaluation of whether we are in fact comparing two comparable operations, and whether we have correctly deduced which are the most primitive operations. Even supposing we have correctly determined which operations are primitive, there remains the matter of demonstrating empirically that the grammar in fact favours primitive operations over more complex (or composite) ones.

3.3 Critical responses to Merge over Move

There has been much succeeding literature that has criticized, or at least called into question the evidence for, the Merge over Move principle, including Shima (2000), Castillo et al (1999), Richards (1999), Groat (1999). Johnson and Lappin (1997) argue against MOM resulting in a

---

9 There is another theory-internal argument for Move being more complex than Merge: that Merge is theoretically more necessary in some way to any grammar (Chomsky 1995 calls this “virtual conceptual necessity”).
reduction of computational complexity. Castillo et al (1999) present both theoretical and empirical arguments against the principle, as does Shima (2000). Even authors who do not take a strong stance against the principle (indeed, even those who seem to support it) remark on its lack of empirical support. Hornstein (2009: 50) notes that “the empirical evidence that MOM is a principle of grammar is not overwhelming,” and Boeckx and Grohmann (2007: 210) state, “We cannot fail to notice that in recent years, MOM has become quite suspect.” The reasons for this skepticism are both theoretical and empirical, stemming from a “dramatically changed” conception of Move (Boeckx and Grohmann 2007: 210) since the inception of the principle, as well as a rash of alternative analyses which have sprung up for the core empirical case on which the principle rests. The following sections deal first with responses to the theory-internal motivation for the principle, and second, with the empirical evidence that was first proposed by Chomsky to motivate this principle.

Castillo et al (1999) present a pertinent discussion of the inconsistency of Chomsky’s (1995, *inter alia*) formulation of MOM. They note that MOM is formulated to decide locally, at a particular step in a derivation, which of two operations is more economical, but that the principle also must take into consideration the endpoints of each possible derivational path which branches from that local decision-point. They state that this is problematic: “convergence is by its very nature a property of derivations stated over outputs, whereas the comparison metric must decide locally.” Chomsky’s (1995: 220-221) assumption that economy principles like MOM act on sets of convergent derivations (a ‘global’ quality) is grounded conceptually in the following idea: the most economical derivation one can imagine is one in which no operations apply at all. Chomsky (1995: 221) states that such a derivation would surely crash, thus economy conditions must be global:

“If nonconvergent derivations can block others, this derivation [the one in which no operations at all apply] will block all others and some elaboration will be needed […]. In the absence of convincing evidence to the contrary, then, I will continue to assume that economy considerations hold only of convergent derivations: D_{\Lambda} is a subset of D_{C}.”

Castillo et al (1999: 4) argue that the derivational paths that lead to (6a) and (6b) are both equally costly from a global perspective since they both involve one operation of overt movement; furthermore, they both converge, indicating the importance of a local evaluation of what the ‘cheapest’ next step is. (In other words, a strictly ‘global’ MOM principle, which only pays attention to outputs, is untenable.) Nevertheless, a strictly local interpretation of MOM is equally problematic: the principle must not operate over derivations that do not converge, not only for the general conceptual idea in the above quotation, but for hard empirical reasons. The contrast in (19) (example from Castillo et al 1999: 5) indicates that a strictly local version of MOM will make the wrong empirical predictions. If we Merge the expletive *it* at the state in (20), the derivation will eventually crash for violating the Minimal Link Condition; if we Move *John*, on the other hand, the derivation will converge at (19b).

(19)  
  a. *John seems that it was in the room.
  b. It seems that John was in the room.

(20)  
  [TP T [was [John [in the room]]]]

---

10 For further discussion of computational complexity issues and Merge over Move, see Chomsky (1998), Collins (1997), Frampton & Guttman (1998), and Yang (1997). (List excerpted from Castillo et al 1999.)
Castillo et al (1999) argue that “the fact that [19a] crashes cannot be relevant for the contrast in [19] if MOM is forced to act locally to account for [6a,b].” Groat (1999) also notes the ‘mixed’ global/local nature of the principle. Dealing with the confusing nature of this global/local principle has led to similar issues as those raised above in other empirical arguments made both for and against the principle.

3.3.1 Responses to the Proper-subpart argument

There have been several criticisms of and responses to the proper sub-part argument. Boeckx & Grohmann (2007: 210) note that, “in more recent theorizing, ‘Move’ is just one of the forms of the basic operation Merge.” These authors point out that later Chomsky (2004, ‘Beyond Explanatory Adequacy’) presents internal Merge (what we originally would have called ‘Move’) as a subtype of the Merge operation more generally. Chomsky (2004) erases the distinction between internal and external merge, so that the only difference between the operation which selects from the numeration and merges an element into the tree, and the operation that selects a sub-component of the tree and re-merges in into the tree, is the component being merged. Boeckx and Grohmann (2007: 210) argue that, “under this conception of Merge and ‘Move,’ it becomes less clear how the economy-based conceptual argument motivates MOM. Specifically, internal Merge (‘Move’), as defined by Chomsky in the above quote, does not appear to be more complex than external Merge any longer. Thus, Merge-over-Move loses its conceptual argument.” It appears that subsequent revisions to the conceptualization of “Move” have nullified the grounds on which Merge was at one point argued to be less complex than Move.

Shima (2000) also presents an argument against the Proper-subpart argument for Merge over Move. His argument is that Merge involves an additional step that Move does not: “[the selection of] a lexical item from the initial numeration and its introduction into the derivation.” Shima identifies this operation of selection as being, specifically, Select from Chomsky (1995: 226). This Select is thus implied to be different from the operation of select which is a subcomponent of Move ([b] in the above block quotation from Shima 2000). The selecting component of Move is, specifically, the selection of a phrase already within the derivation, on the basis of its having a feature F. It does seem reasonable to suppose that these are two different operations of selection, since one is selecting from the numeration (either ‘blindly’ or based on whatever determines selection from the numeration) and the other is selecting from within the structure already partially assembled, based on a feature-specific trigger. This is the basis of Shima’s comment that this is “a step not involved in applications of Move.” Specifically, Shima (2000: 376) argues that once we have reached the crucial point in the derivation of the two sentences in (6), where the two derivations diverge depending on whether one merges there or raises someone, there is no difference in the number of (operational) steps needed to insert there from the numeration or to raise someone, thus nullifying the argument that Merge is less costly in this operational sense:

“The insertion of there into [Spec, TP] involves three subcomponents: (a) selection of there from the numeration and its introduction into the derivation, (b) merger of there and TP, and (c) agreement between a feature of there and the corresponding feature of T\(^0\). The raising of someone also involves three subcomponents: (a) agreement between a feature of someone and the corresponding feature of T\(^0\), (b) pied-piping of someone, and (c) merger of someone and TP.”

One could make a similar argument using Copy as one of the essentially primitive subcomponents of Move. Hornstein et al (2005, and subsequent) argue that once we have accepted the Copy theory of movement for independent reasons, Move decomposes into Copy + Merge, and is
therefore more complex (involving two primitive or atomic operations rather than one). If we consider the operation Select (as “Select” from the numeration), then Merge and Move both involve the same number of primitive operations, (two). Move is then Copy + Merge. But Merge is then Select + Merge. The two operations are equal in computational complexity, and thus we would not expect Merge to be preferred.\(^\text{11}\)

It is difficult to decide which should be our primitive operations. Merge certainly seems to be well motivated, but if empirical phenomena can be explained by decomposing Merge into small operations (as in Hornstein & Nunes 2009), then we must determine which operations ought to be counted as steps when evaluating economy. Furthermore, what evidence do we have that the primitive operations are all of equal computational complexity, an assumption implicit in the argument from operational complexity as economy? Answering these questions is crucial to building an argument like the proper-subpart argument, so while we can speculate that Merge might be primitive, while Move is tripartite, we should also think about what kind of evidence, theory-internal or empirical, supports the assumption that primitive operations are all equally costly, and what evidence supports identifying these particular operations as primitive.

It is easy to imagine that the costliness of individual, atomic (primitive) operations could be based on other distinctions among them. For example, Shima (2000) suggests that a conceptual reason for preferring Move over Merge is that Merge combines components which are from separate structures, while Move combines components that are within a single structure. Shima argues that, “in the process of forming phrase structures, it is more economical to look only at an already formed structure than to look at, not only an existing structure, but also lexical items in the numeration, or at an independent syntactic object” (p. 376). From purely conceptual (or theory-internal) grounds, the evidence for which operations are more ‘economical,’ if such a distinction exists at all, remains an open question. From conceptual arguments alone, it is not apparent which metric of economy is more relevant to determining operational complexity.

Deal (2009:21) has also constructed an empirical argument that, “economy constraints on derivations actually favour movement (work fewer resources harder) over merger (spend, spend, spend!)”.\(^\text{12}\) Implicit in this conception of Merge seems to be the idea that Merge is not ‘costless,’ as was once suggested. The original impetus (Chomsky 1993) for Merge being costless was that Merge was required by “virtual conceptual necessity.” It is far from clear whether such ‘virtual conceptual necessity’ is enough justification for a kind of operational simplicity within the grammar.

Several things need to be determined in order to give the proper-subpart argument legitimate conceptual validity: \(^\text{13}\) does the grammar compare the operations it has at its disposal at every step? 2) What makes two operations comparable? 3) If two comparable operations are compared at a step in the derivation, does the less complex operation pre-empt the more complex one? 4) And, in the above case, what defines ‘complexity’? The proper-subpart argument is committed to a specific

\(^{11}\) I can think of a counter-argument here, which goes something like this: if we assume movement involves some kind of selection as well, not from the numeration but from the structure already in the workspace—perhaps then equivalent to the selection operation which is identified as a subcomponent of Move in Chomsky (1998) and Shima (2000)—then Merge is really Select + Merge, and Move is really Select + Copy + Merge; therefore, Move continues to be operationally more complex than Merge, and Merge is preferred for economy purposes. I think the reasons for putting this argument aside include the fact that this analysis does not consider the role of AGREE, as Chomsky (1998) and Shima (2000) do; furthermore, the fact that counter-arguments and counter-counter-arguments to the Proper-subpart Argument can be so easily constructed is evidence that we really cannot be sure what is the best way to ‘count’ primitive operations, or even to decide which operations should be primitive. Hornstein & Nunes (2008) have suggested that Merge is really composed of two yet more primitive operations, Concatenate and Label. Whether there are implications of this hypothesis for the Proper Sub-part Argument has yet to be shown.

\(^{12}\) See Deal (2009: 21) for details.

\(^{13}\) The assessment here focuses almost entirely on the ‘local’ side of MOM.
answer on each of these four points, without independent conceptual or empirical evidence for each point. In sum, the proper-subpart argument fails to be convincing because it rests on assumptions that are too many and too interdependent, without adequate empirical support.

3.3.2 A Conceptual Argument against MOVE = MERGE + COPY

The proper-subpart argument assumes that Move is composed of Merge and Copy. It therefore follows that Merge is ‘smaller’ or ‘less complex’ than Move. It is also assumed in this argument that we are dealing with the same Merge operation, or its ‘complexity’ wouldn’t necessarily be the same from one version of it to the other (internal versus external). This Merge operation is an autonomous operation, in the sense that it can operate on its own—it is not required to co-occur with another operation.

It can be readily established that Copy cannot be autonomous—it generally cannot happen independent of all other operations (see Hornstein 2001: 99-102 for discussion). The reason is that an autonomous Copy leads to rampant overgeneration. Copying components of the derivation and then leaving them around to be merged whenever is convenient (rather than immediately following the application of Copy) leads to the derivation of structures that are not grammatical. Copy is not, therefore, a primitive operation. But it is hard to see how one can motivate Move as being non-primitive—that is, composed of certain subparts—when at least one of these subparts does not (indeed, cannot) occur independently. The question then becomes whether an operation can be a ‘proper subpart’ without being an independent operation: another tricky and perplexing question which cannot simply be glossed over.

Thus, the somewhat standard idea that Move is not primitive—that it is actually composed of Copy + Merge (or something similar)—runs into the following difficulty: this conceptually raises the question of whether Copy can be a proper subpart without being autonomous. One could argue that, if Copy is not autonomous, Move cannot be decomposed into its subparts, and is therefore in some sense primitive. If Move is primitive, then the Merge over Move story based on operational complexity falls apart.

Hornstein (2001: 99-102) presents an argument that the “component operation,” Copy, “must be constrained” or we end up with massive overgeneration in the grammar. He proposes that there must be some kind of condition on applications of Copy that restrict it to copying an element only when that element is about to be re-merged. He proposes the following principles:

(21) A copy C made at step N of a derivation must be grammatically integrated at step N + 1.

---

14 This assumes that in order for an operation to be ‘primitive’ it must also be independent/autonomous; this is certainly not an uncontroversial claim.

15 Hornstein (2001) argues in a footnote that Copy actually does occur not coupled with Merge in his analysis of certain sentences, but I leave the detailed examination of these examples to later research.

16 Such overgeneration is empirically demonstrable, and follows from copying components and then leaving them around in the ‘workspace’ to be merged at a later point—essentially, letting Copy act as (what I’ve been calling) an ‘autonomous’ operation. Hornstein provides the following example of an ungrammatical sentence that could be formed assuming an unconstrained Copy:

(i) *Which book did you read Moby Dick before Frank reviewed (Hornstein 2001:100)

Using an unconstrained Copy operation, this ungrammatical sentence is formed as follows (from Hornstein 2001:100): “Form the adjunct. Make a copy of ‘which book’ but do not merge it anywhere. Then build the matrix VP […] adjoin the adjunct to this VP. Then build IP and move ‘you’ to its spec. Continue merging functional categories and build the CP. Finally, merge the copy of ‘which book’ made earlier […] into the Spec CP. This derivation obeys extension and greed at every point.”
(22) An expression $E$ is grammatically integrated iff $\text{def } E$ is a proper subset of a phrase marker.\textsuperscript{17}

The result of this is that Copy becomes ‘bound’ again, in some sense, to Merge in the context of Move. If such stipulations are necessary in order to make Copy viable as an independent operation, one might wonder why it is so important that Copy be an independent operation, since its application seems so closely bound to instances of internal Merge.\textsuperscript{18} It might be that there is little else to recommend the Move = Copy + Merge story without Merge over Move.

It is not even particularly clear what benefit, aside from the support for MOM, accrues to us if Move really is composed of two independent sub-operations. Hornstein (2001: 101) highlights one particular problem: “If [as (21) above stipulates] Merge is delayed then how is it that at a given point of the derivation one evaluates whether it is better to merge or to copy at that point? Asked another way: as both Merge and Copy are single operations, why should Merge be better than Copy if all one does is count operations?” Hornstein is pointing out that, if Copy and Merge really are independent operations, at a given stage $N$ where the next operation could be either Copy or (external) Merge, the grammar is not really comparing two operations of different degrees of complexity, but rather two operations of equal complexity. The Merge-over-Move argument based on operational complexity has disappeared. It would seem that adopting a Move = Copy + Merge hypothesis in order to explain how Move is more complex than Merge leads to the failure of the original principle that inspired this account, since now we are comparing Merge to Copy, and Copy alone.

Hornstein (2001: 102) has a response to this. He encourages us to assume that “operations that result in feature checking trump those that do not”. If feature checking has reached a halt without applying Copy (e.g., with successive applications of Merge), and applying Copy at this stage will result (eventually) in convergence, then Copy will apply. This makes the prediction that (*pure*/‘external’) Merge will always trump Copy unless further (external) Merge operations will not result in convergence, while Copy (followed by internal Merge) will. This relies, obviously, on a fairly robust look-ahead ability in the grammar. But even if we are willing to put aside our reservations about the look-ahead, what has this really done for the proper-subpart argument, based on operation complexity? Under this account, we are still only comparing the operation Copy to the operation (external) Merge. The grammar is no longer choosing Merge over Move because Merge is less complex; instead, it is choosing between Copy and Merge, which, at least under the assumptions held so far, are of equal complexity, both being ‘primitive.’ There is no reason given, then, why Merge should be the default until look-ahead reports back to the derivation that an application of Merge will result in non-convergence, while an application of Copy will subsequently result in convergence.\textsuperscript{19} The conceptual argument for why Merge is simpler than Move has disintegrated. The new argument that is left in its place is simply that Merge is (stipulated) the default choice of operation when a choice is available at a particular point in a derivation, and therefore Merge will be

\textsuperscript{17} I will not directly address here the fact that, as stated, these are stipulations. Hornstein does argue that these might follow from more general principles (specifically, operations leading to convergence might be favoured over operations that do not lead to convergence), but I do not find his specific arguments compelling.

\textsuperscript{18} My observations here are similar to J.C. Castillo’s, as reported in Hornstein (2001: 101, fn. 61): “… Castillo observes that this seems very close to reintroducing the operation Move as a primitive: why should the two operations be required to work so close together?”

\textsuperscript{19} E. Cowper notes that there may be conceptual motivation for Merge to be the default if every instance of Copy also triggers an application of Merge; then, Copy automatically adds more steps to the derivation. However, this assumes a) that we are counting steps, and b), that Copy and Merge each count as one full step. The theoretical motivation for counting steps is generally weak; further conceptual justification and empirical motivation is needed to accept either a) or b).
chosen before Copy until a point where the only future for the derivation via Merge is non-convergence. If, at this point, there is a convergent path the derivation can take that takes Copy to be its next step, then Copy will apply, followed next by (internal) Merge. But this only leaves us with several stipulations, rather than a conceptual reason for why Merge should be the default operation.

As for the criticism that the conditions in (21) and (22) above merely re-introduce the idea of a primitive Move (observed by both J.C. Castillo and me), Hornstein’s response is that Merge and Copy are not separate operations, but “sub-parts of a single operation.” This amounts to having the best of two incompatible worlds: Move must be a composite operation to make the proper-subpart argument work, but it is difficult to account for the status of Copy without appealing to the singularity of the operation Move. In order to make the Move = Merge + Copy argument for MOM work, it seems necessary to conceive of Move as being both a single operation and two separate operations. It is thus not clear to me what a ‘sub-operation’ is, on a complexity scale, in relation to other operations which are not ‘sub-operations’ but full ones. If we cannot define what it means to be a subpart, why then go on to make conceptual claims based on the property of ‘being a subpart’? We might expect this kind of paradox, that Move is both a primitive operation and a composite one, to be the sort of wall one runs up against when the paradigm within which one currently is working is not sufficient to solve the problem at hand. We do not have a clear enough understanding of what makes an argument primitive, and what makes primitive arguments comparable in complexity terms.

3.3.3 Empirical Responses to MOM

Moving past the conceptual arguments for Merge over Move, the empirical evidence for MOM as an economy principle is scant. Not only can empirical arguments be found against the principle (Shima 2000, Deal 2009: 21), but the lack of empirical evidence for the principle is also worrisome. This distinction points to two kinds of empirical evidence against Merge over Move: empirical cases where Merge over Move predicts the wrong result, and alternative analyses that solve the conundrum of the data in (6) without recourse to Merge over Move. The first kind of empirical cases are discussed in 3.3.3.1 below; the second are dealt with at length in section 3.4.

3.3.3.1 Empirical Arguments against MOM

Shima (2000) presents two empirical arguments, both with English data, against the MOM principle by demonstrating empirical motivation for an economy principle which is the exact opposite: Move over Merge. Shima’s arguments here are not only empirical cases against Merge over Move, but empirical arguments for another economy principle, Move over Merge. We will first deal with this argument before moving on to alternative analyses of English expletives, since empirical arguments against Merge over Move which simply replace it with another economy principle would return us to the central question of this paper: whether economy principles like Merge over Move (or Move over Merge) are actually empirically necessary components of UG. Further potential empirical cases against the principle will present themselves in the course of the discussion of alternative analyses of (6) presented in section 3.4.

3.3.3.1.1 Shima (2000): Locality of A-Movement

Shima (2000) labels his two empirical arguments for Move over Merge (contra Merge over Move) the ‘Locality of A-movement’ argument and the ‘Strict cyclicity’ argument. The first of these is based primarily on the ungrammaticality of the sentence in (23) (and [44], [45]) below. According to Shima, this ungrammatical sentence is derived in the Chomskian framework assuming Merge over Move, while it is not derivable (as it should be) within the same framework, but using Shima’s
proposed Move over Merge economy principle instead of MOM. The argument relies on Shima’s rejection of defective intervention constraints\(^{20}\) in favour of a rather stipulative condition on Case.

(23)  *John seems that it is likely John to win.

To begin with a sub-part of Shima’s argument, Shima rejects defective intervention constraints based on the data in (24) (and, later, [34a]) below. The derivation of this sentence proceeds as follows in (25) – (32) until a decision point is reached at (32).

(24)  They seem to him to like John.

(25)    a.  \[N_0 = \{\text{John}_1, \text{like}_1, \text{to(T)}_0, \text{T}^0_1, \text{they}_1, \text{seem}_1, \text{to(P)}_0, \text{him}_1, v^0_2\}\]

(26)    a.  \[A = [\text{TP} \text{ to } \{\text{they}_1 v^0 \text{[VP like John]}\}]\]
   b.  \[N_A = \{\text{John}_0, \text{like}_0, \text{to(T)}_0, \text{T}^0_1, \text{they}_0, \text{seem}_1, \text{to(P)}_0, \text{him}_1, v^0_1\}\]

(27)    a.  \[B = [\text{TP} \text{ they } \{\text{they}_1 v^0 \text{ like John}\}]\]
   b.  \[N_B = \{\text{John}_0, \text{like}_0, \text{to(T)}_0, \text{T}^0_1, \text{they}_0, \text{seem}_1, \text{to(P)}_0, \text{him}_1, v^0_1\}\]

(28)    a.  \[C = [\text{V} \text{ seem } \{\text{TP they } \{\text{they}_1 v^0 \text{ like John}\}]\]
   b.  \[N_C = \{\text{John}_0, \text{like}_0, \text{to(T)}_0, \text{T}^0_1, \text{they}_0, \text{seem}_1, \text{to(P)}_0, \text{him}_1, v^0_1\}\]

(29)    a.  \[D = [\text{VP } \text{ PP to him } \{\text{V} \text{ seem } \{\text{TP they } \{\text{they}_1 v^0 \text{ like John}\}\}\]
   b.  \[N_D = \{\text{John}_0, \text{like}_0, \text{to(T)}_0, \text{T}^0_1, \text{they}_0, \text{seem}_1, \text{to(P)}_0, \text{him}_0, v^0_1\}\]

(30)    a.  \[E = [\text{VP } \text{ PP to him } \{\text{V} \text{ seem } \{\text{TP they } \{\text{they}_1 v^0 \text{ like John}\}\}\]
   b.  \[N_E = \{\text{John}_0, \text{like}_0, \text{to(T)}_0, \text{T}^0_1, \text{they}_0, \text{seem}_1, \text{to(P)}_0, \text{him}_0, v^0_1\}\]

(31)    a.  \[F = [\text{VP } \text{ v}^0 + \text{ seem } \{\text{VP } \text{ PP to him } \{\text{VP } \text{ seem } \{\text{TP they } \{\text{they}_1 v^0 \text{ like John}\}\}\}
   b.  \[N_F = \{\text{John}_0, \text{like}_0, \text{to(T)}_0, \text{T}^0_1, \text{they}_0, \text{seem}_1, \text{to(P)}_0, \text{him}_0, v^0_1\}\]

(32)    a.  \[G = [\text{TP } \text{ T}^0_0 + \text{ seem } \{\text{VP } \text{ PP to him } \{\text{V} \text{ seem } \{\text{TP they } \{\text{they}_1 v^0 \text{ like John}\}\}\}\]
   b.  \[N_G = \{\text{John}_0, \text{like}_0, \text{to(T)}_0, \text{T}^0_0, \text{they}_0, \text{seem}_1, \text{to(P)}_0, \text{him}_0, v^0_1\}\]

\([32a] \text{ directly above corresponds to Shima’s [9].} \) The ‘decision’ to be made at (32) is whether to move him, or they, up to fill [Spec, TP] of the matrix TP. If defective intervention constraints are correct, then him, closer to matrix T\(^0\) and still with \(\varphi\)-features visible (though case has been checked), will block they from moving to fill [Spec, TP]\(\text{matrix}\). This is not the case, since the sentence in (24) is in fact grammatical, and the final stage of the derivation looks like this:

(33)    a.  \[H = [\text{TP they } \{\text{T} \text{ T}^0_0 + \text{ seem } \{\text{VP } \text{ PP to him } \{\text{V} \text{ seem } \{\text{TP they } \{\text{they}_1 v^0 \text{ like John}\}\}\}\]\]

\(^{20}\) The defective intervention constraint (Chomsky 2000:123) states that, in the schematic (i) below, when \(\alpha\) is a probe and \(\beta\) and \(\gamma\) are both matching goals, \(\alpha\) is blocked from Agree with \(\gamma\) by the presence of \(\beta\), even if \(\beta\)’s features have already been checked (by an earlier Agree operation) and \(\beta\) is therefore inactive.

(i.)  \(\alpha > \beta > \gamma\)
There is, according to Shima, further evidence against defective intervention constraints in sentences of the type in (34a) below. Without going through the entire derivation, take it for granted that at a certain point the derivation reaches the stage in (34b). According to Shima, if defective intervention constraints were correct, (34b) below would predict the ungrammaticality of the sentence in (34a). In (34b), the raising of this claim is predicted to be blocked by the visible $\varphi$ features of the pronoun in $[\text{Spec}, \text{VP}]$.

(34)  

a. This claim strikes me as correct.

b. $[\text{TP} \ T^0 \ [v^0 + \text{strikes}] [v^0 \text{me} [v^0 \text{strikes} [\text{sc this claim as correct}]]]]$

Shima therefore rejects the idea of defective intervention constraints in favour of the following alternative condition, which allows him then to derive the grammatical sentences in (24) and (34a) above:

(35) $[\text{Spec}, \text{TP}]$ can be filled only by a DP with structural case.

Assuming (35), the grammaticality of (24) and (34a) are predicted in the following way. In (24), the pronoun $\text{him}$ receives oblique case from the preposition $\text{to}$, and thus is not available to fill the position of $[\text{Spec}, \text{TP}]_{\text{matrix}}$, since oblique case is inherent, not structural. Presumably, a derivation which chooses to raise $\text{him}$ will then crash. Only $\text{they}$, having nominative (a structural) case, is predicted to be able to fill $[\text{Spec}, \text{TP}]_{\text{matrix}}$. This is assuming that we have discarded defective intervention constraints, since there is nothing in condition (35) that would necessarily prevent defective intervention constraints from still being a part of the grammar. In order to account for cross-linguistic facts, Shima must also postulate that (35) is a parameter, since in a language like French, a sentence like (24) is ungrammatical, rather than grammatical, as it is in English:

(36) $\text{Jean semble à Marie [jean avoir du talent].}$

Jean seems to Marie to have talent

‘Jean seems to Marie to have talent.’

As for sentence (34a) above, though Shima does not spell out how this case is resolved with the adoption of condition (35) and the rejection of defective intervention constraints, we can see that with these stipulations in place, $[\text{Spec}, \text{TP}]_{\text{matrix}}$ of (34b) must be filled by a DP with structural case. The pronoun $\text{me}$ already has accusative case (from $v^0$), but the DP $\text{this claim}$ does not, and is available to receive nominative (structural) case by raising to the specifier position of the matrix $T^0$. $\text{Me}$ is not available to raise to $[\text{Spec}, \text{TP}]_{\text{matrix}}$ since it has already checked its case, and there is no defective intervention constraint to allow $\text{me}$ to interfere with $T^0_{\text{matrix}}$ probing to $\text{this claim}$. This concludes the sub-argument for the rejection of defective intervention constraints and the adoption of the condition in (35).

We now return to the data in (23), repeated below in (37). The derivation proceeds to the point in (37b), with the numeration as in (37c). According to Shima, once the derivation has reached this point, there are two possibilities: either Merge $\text{it}$ into $[\text{Spec}, \text{TP}]$ or move $\text{John}$ from below up into the specifier of $T^0$. Thus (37b) is another decision point for MOM.

(37)  

a. $\text{*John seems that it is likely John to win.}$

b. $[\text{TP} \ T^0 \ [v^0 \text{is likely} \text{John to win}]]$

c. $\{\text{win}_0, \text{to}_0, \text{John}_0, \text{likely}_0, \text{is}_0, \text{it}_1, \text{that}_1, \text{seems}_1, T^0_1\}$
MERGE OVER MOVE AND THE EMPIRICAL FORCE OF ECONOMY IN MINIMALISM

(38) a. \[[TP \text{it} [T\cdot T^0 [VP \text{is likely John to win}]]\]
   b. \{win_0, to_0, John_0, likely_0, is_0, it_0, that_1, seems_1, T^0_1\}

(39) a. \[[TP T [VP \text{seems} [CP \text{that} [TP \text{it} [T\cdot T^0 [VP \text{is likely John to win}]]]]]]\]
   b. \{win_0, to_0, John_0, likely_0, is_0, it_0, that_0, seems_0, T^0_0\}

(40) a. \[[TP John [T\cdot T [VP \text{seems} [CP \text{that} [TP \text{it} [T\cdot T^0 [VP \text{is likely John to win}]]]]]]\]
   b. \*‘John seems that it is likely to win.’

If we assume MOM, MOM will choose the derivational path that merges it rather than raising John from the embedded clause, leading to the structure in (38). The derivation then proceeds with further merges until reaching the stage in (39). At this point, nothing further is available in the numeration to be merged ([39b]), so John will be raised, deriving the sentence in (40). Shima (2000: 378) writes, “Since it has its nominative Case feature checked in Spec of the embedded TP, [35] allows Spec of the matrix TP to be filled by John with nominative Case rather than it”. So, we have a problem, since (40b) is clearly ungrammatical.

Shima argues that his alternative economy condition, Move over Merge, in conjunction with the stipulated condition (35), will correctly exclude the above sentence. If at the stage in (37b), Move over Merge prompts raising of John instead of merging of it, then we will end up with the structure in (41). Further merges create the structure in (42), at which point it is still in the numeration.

(41) \[[TP John [TP T^0 [VP \text{is likely John to win}]]\]

(42) \[[TP T [VP \text{seems} [CP \text{that} [TP John [TP T^0 [VP \text{is likely John to win}]]]]]]\]

Shima writes, “Condition [35] does not require this position to be filled by raising John since its nominative Case feature has been checked in Spec of the embedded TP. Thus, it is inserted into Spec of the matrix T” (p. 378). Shima’s principle, Move over Merge, would predict John should raise again (although this would leave it in the numeration and cause a crash), but since John has already had its case feature checked in Spec of the embedded TP, John won’t raise (and it is then inserted). The result is that we end up with the grammatical sentence in (43):

(43) a. \[[TP \text{it} [T\cdot T [VP \text{seems} [CP \text{that} [TP John [TP T^0 [VP \text{is likely John to win}]]]]]]\]
 b. ‘It seems that John is likely to win.’

It is important to note how Shima’s Move over Merge story relies on the rejection of defective interveners. If defective interveners are correct, then superraising is quite simply barred by the intervention of it, if it merges low (in the embedded TP), and the argument from economy disappears. Shima’s removal of defective intervention constraints is crucial to his argument, and the evidence on which he excludes them from the grammar is sparse. In particular, one could make the argument that him (in 24/33) in the prepositional phrase [to him] is not c-commanded by the probe (matrix T), particularly if to him were assumed to be an adjunct.

In addition to these doubts about the arguments against defective intervention constraints, we can also criticize the principle proposed in their place, condition (35), for its extremely stipulative nature. Condition (35) amounts to little more than a descriptively adequate condition with little or no theoretical support. In particular, we might ask how the grammar is to know that John will eventually be realized with structural case in the derivation, before it has been raised. This condition has little explanatory value and therefore should be treated with skepticism.
Shima provides two more arguments for Move over Merge, and against Merge over Move, based on ungrammatical sentences that he argues are derivable assuming MOM, and not derivable assuming Move over Merge. The sentences are provided below in (44) – (45).

(44) a. *John is asked [{CP [how likely \textsubscript{1} John to win]} \textsubscript{2}] it is \textsubscript{1} 
   b. \textsubscript{[TP T \textsubscript{6} [VP is [how likely John to win]]]}
   c. \textsubscript{[TP it [T\textsubscript{4} T \textsubscript{6} [VP is [how likely John to win]]]}
   d. \textsubscript{[CP [TP it [T\textsubscript{4} T \textsubscript{6} [VP is [how likely John to win]]]]]}
   e. \textsubscript{[CP [how likely John to win] \textsubscript{6} [C\textsubscript{1} C \textsubscript{6} [TP it [T\textsubscript{4} T \textsubscript{6} [VP is [how likely John to win]]]]]]}
   f. \textsubscript{[TP T \textsubscript{6} is [asked [CP [how likely John to win] \textsubscript{6} [C\textsubscript{1} C \textsubscript{6} [TP it [T\textsubscript{4} T \textsubscript{6} [VP is [how likely John to win]]]]]]]}
   g. \textsubscript{[TP John [T\textsubscript{4} T \textsubscript{6} is [asked [CP [how likely John to win] \textsubscript{6} [C\textsubscript{1} C \textsubscript{6} [TP it [T\textsubscript{4} T \textsubscript{6} [VP is [how likely John to win]]]]]]]}}

(45) *Who\textsubscript{1} was questioned [{CP t\textsubscript{1} [C\textsubscript{1} C <\textsubscript{wh}> [ip it was told t\textsubscript{1} that Mary left]]]?  

To begin with (44), Shima assumes the structure in (44b) for the embedded clause. According to MOM, it is inserted in the [Spec TP] of the embedded clause. Further merge results in the structure in (44d). Then, the complement of is raises to [Spec CP] of the embedded clause (44e). Further merge yields the structure in (44f). Shima then suggests that John raise from the complex specifier of the embedded CP to the matrix [Spec TP] (44g), and that this movement is not barred by any principle or condition within the Minimalist framework, thus yielding an ungrammatical sentence.

However, one criticism that Shima (2000) acknowledges is that the ungrammatical sentence in (44) might be ruled out for its violation of the Condition on Extraction Domain (CED), (Huang 1982), which bars extraction from a constituent that is not governed. Shima’s (2000: 379 fn. 4) response to this is that [how likely (John) to win] is not a barrier to the extraction of John in (44d) because it is the specifier of the complement of the verb asked. Under the assumptions of Chomsky and Lasnik (1993), “XP is not a barrier if it is the complement of a head H or the Spec of the complement of H” (p. 397). Thus, if asked is the head H, then the CP it selects is the complement of H, and the phrase [how likely John to win] is the specifier of the complement of H. Therefore [how likely John to win] is not a barrier to extraction of John.

At first glance, it might seem possible to propose, alternatively to the economy story, that we are unable to extract out of the degree phrase, [how likely John to win], because the degree phrase is some kind of phase, assuming Spell-out by phase. However, data such as that in (46) rules out this over-simple explanation. In a derivation like (46), we must be able to extract John from the degree phrase, providing evidence against the idea that the degree phrase could itself be a phase.

(46) a. \textsubscript{[T\textsubscript{4} [VP is [DegP how [AP likely [ip John to win]]]]]}\textsuperscript{21}
   b. \textsubscript{[TP John [T\textsubscript{4} [VP is [DegP how [AP likely [ip John to win]]]]]}
   c. John is how likely to win?’

Nor does it appear possible to rule out the derivation in (44) by a different derivation-by-phase explanation: that by removing John from the CP, we extract an element that has already been spelled out as part of the larger CP phase. This is because John is in the specifier of the CP. Under standard assumptions, what gets spelled out at the CP-level phase is the complement of C, not what is in its specifier. The specifier of CP is thus a kind of ‘escape hatch’ for movement out of the phase, and will not help us independently rule out the derivation in question.

\textsuperscript{21} Thanks to E. Cowper for pointing out this example to me.
Taking a closer look, though, at the differences between (44) and (46), we can see that the degree phrase remains in object position in (46), but not in (44). We can exploit this to advantage given a more recent theoretical perspective on traditional CED effects by Nunes and Uriagereka (2007: 337 - 342). The authors have shown that CED effects might be derivable from the Linear Correspondence Axiom of Kayne (1994).

(47) **Linear Correspondence Axiom:** A lexical item \( \alpha \) precedes a lexical item \( \beta \) iff \( \alpha \) asymmetrically c-commands \( \beta \).  

(Nunes and Uriagereka 2007: 338)

Nunes and Uriagereka (2007: 338) state that because of the LCA, “a given head can be directly linearized with respect to the lexical items within its complement, but not with respect to the lexical items within its subject [specifier] or adjunct.” They adopt the view that multiple Spell-out is “triggered by linearization considerations” (p. 338), a view also elaborated in Uriagereka (1999). The idea is that complex specifiers and adjuncts must be assembled in a separate derivational workspace (or a separate location of the same workspace), before being merged as a ‘chunk’ into the rest of the tree.

(48) \([XP [VP Y [AP]] [X' X [ZP]]]\)

Given the structure of the tree in (48), the internal components of the complex specifier YP cannot be linearized with respect to elements in the rest of the tree, for failure to establish asymmetrical c-command with these elements. However, these components can be linearized with respect to each other, within the complex specifier (YP). If treated as a single element (ignoring the internal structure of the specifier), then the specifier can also be linearized with respect to other elements of the tree (e.g. the complement, ZP). This suggests that linearization of elements within the complex specifier must occur independently and prior to movement (or at least independently). Uriagereka and Nunes (2007) suggest that the internal structure of a complex specifier is spelled out and linearized in the PF component, independent of the rest of the tree. The complex specifier is then later merged into the rest of the tree. The label of the specifier supplies the “address” for where it belongs in the larger tree. Assuming K is the label of the complex specifier, the authors state, “K itself is still accessible to the computational system, despite the fact that its constituent parts are, in a sense, gone. […] Another way to put it is to say that once the constituent parts of K are gone, the computational system treats it as a lexical item” (p. 339). This could explain why extraction out of [how likely John to win] in (44) is ungrammatical, while extraction out of the same phrase in (46) is grammatical: in (44) the degree phrase is a complex specifier, with its internal components requiring linearization prior to movement, while in (46), the degree phrase is not linearized separately since all internal components of the phrase are linearizable with respect to all other elements of the tree.

It thus appears to be possible, although in need of a little more elaboration, to derive the intended empirical effect in (44) without telling a Merge over Move story. The ungrammaticality of the data Shima presents seems more likely to be due to some universal constraint on movement rather than an economy principle selecting between operations.

Furthermore, we do not need to look too closely at the Move over Merge story for this piece of data if we accept that the “merge” path in (44) is ruled out by some kind of ‘freezing’ of complex specifiers, motivated by linearization requirements. The derivation that chooses to merge it in at step (44c) will crash independent of economy conditions; on the other hand, the derivational path that moves John up at this stage will eventually converge:

(49) a. \([TP John [T [VP is [how likely John to win]]]]\)

b. \([CP [how likely John to win] C [TP John [T [VP is t1]]]]\)
c. \[ TP \text{ it [ is } [VP \text{ asked } [CP \text{ [how likely } John \text{ to win}] [C’} C \text{ TP John [T’ T [VP is t1 ]]][][][]]] \]
d. ‘It is asked how likely to win John is.’

The “merge path” does not need to be excluded with an economy condition like Move over Merge, as Shima proposes, since it independently does not lead to convergence for the reasons discussed above. The “move path,” on the other hand, starting from the phrase structure in (44c), will not result in a violation from the extraction of John, since John is extracted before movement. So one path never converges for non-economy reasons, and the other path leads to convergence. There is no need for an economy principle here to choose between alternative derivations.

Moving on to (45), we find similar problems with Shima’s argument as we did with the other data. Here, briefly, is Shima’s argument based on (45) above, repeated below as (50).

(50) *Who\_\_1 was questioned \[ [CP t\_1 [C’ C <wh> [IP it was told t\_1 that Mary left]]] \]

Shima assumes this sentence reaches the following intermediate step in the derivation:

(51) \[ [TP T was told who \[ [CP that Mary left]]] \]

Again, Shima argues that we are looking at a decision-point for an economy principle, either MOM or Move over Merge. The ungrammaticality (or grammaticality) of the resulting sentence falls out from the choice made between merging an expletive at this point, or moving the NP Mary already in the structure. Let’s say an available expletive in the numeration merges in at this point, resulting in (52a). The derivation then continues to apparent convergence through steps (52b-e) below.

(52) a. \[ [TP T was told who \[ [CP that Mary left]]] \]
b. \[ [CP C [TP it T was told who \[ [CP that Mary left]]] \]
c. \[ [CP who \[ [C’ C [TP it T was told who \[ [CP that Mary left]]] \]
d. \[ [VP was questioned \[ [CP who \[ [C’ C [TP it T was told who \[ [CP that Mary left]]] \]
e. \[ [TP who \[ [T’ T [VP was questioned \[ [CP who \[ [C’ C [TP it T was told who \[ [CP that Mary left]]]] \]
f. *‘Who was questioned it was told that Mary left?’

Shima argues that the above derivation leaves no formal features unchecked, and therefore should be convergent. He also notes (2000: 380, fn. 5) that “moving a wh-phrase out of [Spec, C] is not barred by Last Resort only if that operation leads to establishing a checking relation.”

This is Shima’s last ‘Locality of A-Movement’ argument against the MOM principle. However, as with the other arguments, it appears there are independent ways in which the problematic data in (45)/(52) could be excluded. In particular, with the above derivation, the movement of the wh-phrase out of [Spec, CP] is suspect. Shima mentions the following counter-argument: “one might argue that [52e] is uninterpretable for the reason that a wh-phrase does not provide [Spec, C] with a wh-feature, and that this LF uninte-rpretability makes the derivation crash” (p. 380). His response to this is to appeal to copy theory. According to Shima’s interpretation of copy theory, “[Spec, CP] is filled by who at LF, as follows:

(53) \[ [TP who\_\_1 [T’ T was questioned \[ [CP who\_\_1 [C’ C <wh> [TP it [T’ T was told who\_\_1 [CP that Mary left]]]]]] \]

Note that the only difference between (53) and (52e) is that in the former the chain created by movement of who has been reduced to one copy by deletion of lower copies, while in (53), Shima’s
representation, all copies are apparently still present. This makes the prediction that a single \textit{wh}-element should be able to check multiple \textit{wh}-features in multiple [Spec, C] positions in the same derivation, which is not empirically borne out, as evidenced by the ungrammaticality of (54) below.

(54) *Who did John wonder ate the cake?

The problem with Shima’s argument on this point might also be related to the fact that, at LF as well as PF, chains need to be reduced to one copy: at LF, this amounts to saying that an element cannot be interpreted in more than one location of its chain. In any case, Shima’s interpretation of the copy theory for \textit{wh}-elements will not make the correct empirical predictions.

Putting this issue aside, the movement from an A position, to an A’-position in (52), then back to an A position is clearly illegal. To defend Shima’s point of view, that this improper movement originates in a choice between Merge and Move, one needs to demonstrate that this strong empirical generalization \textit{always} follows from a choice based on the economy condition. Without investigating this question thoroughly, we cannot be convinced that this is an empirical argument for Move over Merge. The alternative is that it is simply a case where one derivation path leads to a crash for economy-independent reasons, and one leads to convergence.

Overall, Shima’s arguments against Merge over Move do not present enough evidence against MOM, or evidence that does not provide extensive enough empirical support for the breadth of the claim. Furthermore, the arguments he has constructed so far for his alternative economy principle, Move over Merge, also do not provide enough empirical and theoretical support. We need to look elsewhere.

3.3.3.1.2 Shima (2000): Strict Cyclicity Argument

In his second empirical argument against MOM, Shima (2000) argues that Move over Merge, but not Merge over Move, “captures the strict cyclicity property of movement” (p. 382), based on sentences like (55) that violate the CED by extraction out of an embedded subject. Shima assumes that at a certain point in the derivation, the phrase marker looks like (56):

(55) *Who did you say that [pictures of t2]1 were stolen t1?

(56) [TP T [\textit{VP} were stolen [pictures of who]]]

Again, Shima takes this to be a point where one can either Merge or Move, to be determined by an economy condition that compares the options. The goal is to correctly derive the ungrammatical sentence in (55). By ‘correctly,’ I mean (following Shima) a cyclic derivation that moves the embedded object [pictures of who] up before extracting \textit{who}. This violates the Condition on Extraction Domain (CED), and Shima assumes that this violation is the correct reason for the sentence to crash. This derivation is outlined in (57a - d) below.

(57) a. [TP T [\textit{VP} were stolen [pictures of who]]]
   b. [TP [pictures of who] [\textit{T'} T [\textit{VP} were stolen [pictures of who]]]]
   c. [TP you [\textit{VP} say [\textit{CP} that [TP [pictures of who] [\textit{T'} T [\textit{VP} were stolen [pictures of who]]]]]]]

22 Note that Shima gives a double question mark [??] as the grammaticality judgment for this sentence, but I think this sentence is more properly awarded an asterisk [*], especially since the point of this derivation, for Shima, is that it is ungrammatical, or at least ill-formed.
d. $[CP \text{ who } [C \cdot C \ [TP \text{ you } [VP \text{ say } [CP \text{ that } [TP \text{ [pictures of who]} \ [T \cdot T \ [\_P \text{ were stolen \ [pictures of who]}]]]]]]]]$

Shima provides an alternative, ‘non-cyclic’ derivation, in (58) which he argues is ruled out by strict cyclicity. It cannot be ruled out by violation of the CED because the extraction of who should be legitimate, as in (59). Instead, it is the movement of the embedded object, [pictures of t1], to the embedded [Spec TP], in a non-cyclic fashion, which is a problem for this derivation. Thus strict cyclicity is a necessary property of the grammar, and is needed in order to rule out the derivation in (58). Without strict cyclicity, Shima argues, the derivation in (58) would be legitimate.

(58) a. $[TP \ T \ [\_P \text{ were stolen pictures of who}]]$
b. $[CP \text{ that } <C> \ [TP \ T \ [\_P \text{ were stolen pictures of who}]]]
c. $[TP \text{ you say } [CP \text{ that } <C> \ [TP \ T \ [\_P \text{ were stolen pictures of who}]]]]$
d. $[C \cdot C \ [TP \text{ you say } [CP \text{ that } <C> \ [TP \ T \ [\_P \text{ were stolen pictures of who}]]]]]
e. $[CP \text{ who } [C \cdot C \ [TP \text{ you say } [CP \text{ that } <C> \ [TP \ T \ [\_P \text{ were stolen pictures of who}]]]]]]$
f. $[CP \text{ who } [C \cdot C \ [TP \text{ you say } [CP \text{ that } <C> \ [TP \text{ [pictures of who]}] \ [T \cdot T \ [\_P \text{ were stolen t1}]]]]]]]

(59) Who1 did you steal a picture of t1?

Having established the importance of strict cyclicity to the grammar, Shima attempts to show that this property follows from Move over Merge. As stated before, the representation in (56) is the decision-point for this economy principle. According to Shima, the two options we compare at the point of the representation in (56) are the following: “merger of C in the numeration and TP [or] raising of the embedded object to [Spec, TP].” For this argument to work, we have to assume that we are looking at these derivations without strict cyclicity. Shima is trying to derive strict cyclicity from Move over Merge, so we have to imagine what happens with each of these paths without strict cyclicity. Shima states that his economy principle, Move over Merge, derives the correct result here, opting to move [pictures of who] up to [Spec, TP], as in (57), rather than merging C. The non-cyclic derivation in (58) follows from choosing (non-cyclic) Merge (of C) instead of movement of the embedded object at the decision point in (56).

Shima may succeed in showing that strict cyclicity follows from picking Move instead of Merge, formulated as above, but only if we are already not assuming strict cyclicity from other independent considerations. Strict cyclicity may follow from Move over Merge (but not MOM) in the case Shima has highlighted, but he does not succeed in showing strict cyclicity always follows from Move over Merge (and never from MOM), and thus he should not succeed in convincing us that Move over Merge is necessary to derive strict cyclicity. A great deal more data must be considered to prove this; we cannot simply assume that the empirical coverage of the Move over Merge principle, and that of the more general idea of strict cyclicity, will perfectly coincide. This seems like too great a leap of faith to take on the basis of a single ungrammatical sentence.

For example, it might be possible to construct an example where two derivational paths, one cyclic, and one non-cyclic, diverge in two different Merge options (e.g. perhaps merging an expletive in an embedded TP or a complementizer). In this case, strict cyclicity could be argued to still be integral to ruling out the non-cyclic derivation, but in this case, where the two possible paths do not have a movement option, it would be impossible to argue that the proposed economy condition Move over Merge derived this strict cyclicity effect.

Furthermore, there is independent evidence for the necessity of strict cyclicity if one starts by assuming some version of multiple spell-out or spell-out by phase. If one is convinced that there is independent empirical evidence for spellout by phase, then strict cyclicity automatically follows,
making Shima’s Move over Merge argument redundant. If we are spelling out ‘chunks’ of the tree as we go, we do not have the opportunity to move the embedded object in (58) non-cyclically after the derivation has progressed beyond the embedded CP stage. The embedded, spelled-out TP (complement of C) will have unchecked EPP and ϕ features, causing a crash.

Further empirical investigation into this issue should demonstrate that strict cyclicity will not always be derivable from Move over Merge, though I leave this proof to further work. The derivation in (58) that Shima presents as evidence against the Merge over Move principle seems untenable from a number of perspectives, none of them directly related to economy.

Strictly speaking, there is also one final way in which Shima’s argument diverges from Chomsky’s original argument: according to Chomsky, an economy principle like Merge over Move selects from among convergent derivations. The set of derivations selected by the economy principle is a subset of convergent derivations. Chomsky (1995: 220-221) states:

“The language L […] generates three relevant sets of computations: the set D of derivations, a subset D_C of convergent derivations of D, and a subset D_A of admissible derivations of D […] the economy conditions select D_A. In chapters 1-3 it was assumed that economy considerations hold only among convergent derivations; if a derivation crashes, it does not block others. Thus, D_A is a subset of D_C.”

Shima’s Move over Merge argument is clearly different, since the derivation that Move over Merge selects is not a convergent derivation, but a derivation that crashes for the right reasons, so to speak (the CED effects). Even though this more local formulation (without the look-ahead that selecting from among convergent derivations entails) may be conceptually desirable, it has the effect of making an economy principle unnecessary. If one derivational path will crash for an independent reason (or if both will), then there is no need for MOM or Move over Merge to intervene and select one of two convergent derivations as the ‘acceptable’ derivation. This is precisely what makes the data in Chomsky’s original empirical argument for MOM particularly challenging: that we have two apparently convergent derivations, one of which clearly is ungrammatical. It is precisely the move away from this formulation of the economy principle that weakens arguments of a more local kind of economy.

3.3.3.2 Empirical Arguments for the core data, sans MOM

Though we can find evidence that MOM makes the wrong predictions empirically, there is far more evidence from the literature for the non-necessity of the principle. Indeed, there is such a collection of alternative analyses for the data in (6) that we can begin to see how their proliferation is itself an argument against the empirical basis of MOM. The rationale for doing away with MOM based on alternative analyses is as follows. The data in (6) appears to be one of the rare cases where MOM is supported empirically. As noted earlier, Hornstein (2009: 50, fn. 70) observes that, “the empirical evidence that MOM is a principle of grammar is not overwhelming.” The significance of there being very little to empirically motivate this principle is different from the significance of instances where it can be argued the MOM does not in fact derive the correct data at all. The importance of the former kind of argument is that having such sparse empirical data to support the principle argues against its inclusion in the principles of UG. If the MOM principle is motivated solely by two (types of) sentences of English (those in [6] above), plus by perhaps a few very rare other cases (to be dealt with later on in this paper), then the burden of proof shifts from

---

demonstrating that the principle is motivated to showing that it is necessary. The way to go about doing this is to see if it is possible to account for the errant empirical facts without resorting to this principle. Such is the task addressed in the following section.

3.4 Alternative analyses of the core empirical case for MOM

In this section, I will discuss various proposals that have been given for the resolution of the contrast in (6), repeated below as (60), without resorting to the economy principle Merge over Move.

(60)  
   a. There seems to be someone here.  
   b. *There seems someone to be here.

Given the nature of the data, analyses have typically adopted one of two general strategies: either they get rid of the EPP, or they do away with the Chomskian AGREE operation (or restrict it to local, rather than long-distance, AGREE). Furthermore, there is an entire class of alternative analyses proposed by various individuals, which are based on assumptions first proposed by Lasnik (1995) and Belletti (1988), so further research on this topic should address these assumptions in detail. In the sections that follow I discuss some of the key proposals for alternative analyses of the original data in (6), and evaluate which of these are the most promising.

3.4.1 Shima (2000): A Case-Based Account

Shima starts with the following crucial assumptions:

(61)  
   a. SpecTP can be filled only by a DP with structural case.  
   b. The expletive there has a Case feature, and a postcopular DP is optionally assigned partitive Case by a copular (Belletti 1988, Lasnik 1995a,b).  
   c. The expletive there has a formal feature to be checked by that of a DP with partitive Case (Lasnik 1995a,b).

Shima also assumes his own economy principle, Move over Merge. Assumption (61a) was employed in the Move over Merge arguments described in previous sections (see [35]). Assumptions (61b) and (61c) are invoked specifically to deal with the expletive data in (6). The optionality in assumption (61b) is operationalized by having two copular be’s, one which assigns partitive case, and one which doesn’t. This accounts for the contrast in the following data, used as independent evidence for assumption (61b) in Shima (2000: 382-383):

(62)  
   a. There is a man in the room.  
   b. A man is in the room.

In (62a) above, a man checks partitive case with the copula be, while in (62b), a man checks nominative case with T0. The u-nominative case on T0 in (62a) is checked by the expletive there. This contradicts Chomsky’s (1995) assumption that expletive there lacks a case feature. Chomsky’s evidence for this is the sentence in (63).

(63)  
   *[TP There seem [CP that [TP a lot of people are intelligent]]]

The derivation of the sentence in (63) is presumed to crash because matrix T0 has uninterpretable nominative case that remains unchecked at LF. This follows if there does not have a case feature.
Thus, for Shima to assume that *there* can check case in his analysis, he must also have an explanation for why the sentence in (63) is ungrammatical. His explanation is that in (63) there is no DP with partitive case that can check the formal feature of expletive *there* (according to assumption (61c) above). Based on these assumptions, Shima proceeds as follows.

Since optional partitive case is assigned depending on which copular *be* is present in the numeration, Shima outlines both possibilities. First, suppose we have a copular *be* which assigns partitive case to *someone*. Partitive case is inherent, so assumption (61a) doesn’t allow [Spec TP] to be filled with *someone* at the stage in (64) below, even though Move over Merge would predict this.

(64)  \[ [\text{TP} \text{ to } [\text{VP be someone here}]] \]

If *someone* does move up to fill [Spec TP] of the embedded clause, then the derivation crashes. Thus, (61a) has the status of a condition in Shima’s analysis: violating it results in a crashed derivation. On the other hand, if expletive *there* is inserted at this stage, not being barred by assumption (61a), it will presumably check the EPP feature of [Spec, TP]. Since nominative case is structural, rather than inherent, *there* with nominative case is allowed to fill the embedded [Spec, TP]. The derivation continues through the steps in (65) to convergence.

(65)  a.  \[ [\text{TP there} [\text{T to } [\text{VP be someone here}]]] \]
    b.  \[ [\text{TP T [VP seems [TP there [T to } [\text{VP be someone here}]]]]] \]
    c.  \[ [\text{TP there} [\text{T T [VP seems [TP there [T to } [\text{VP be someone here}]]]]]]\]
    d.  ‘There seems to be someone here.’

Here I have a query: if Move over Merge is being postulated as a purely local condition\(^{24}\), then why at stage (64) does applying the principle locally cause a crash? If the principle applies, and *someone* moves, the derivation crashes for violating condition (61a), as described above. If the principle is irrelevant because the outcome is already ungrammatical by virtue of violating condition (61a) above, then the condition must have look-ahead properties and therefore cannot be purely local. The locality of this proposed Move over Merge principle is thus not explicit, and somewhat confused.

To return to Shima’s analysis, suppose we are dealing with the other *be*, the one that does not assign partitive case. Then *someone* (Shima assumes) has nominative case. It can then move up to fill the embedded [Spec, TP] in (64) above; in fact, Shima argues that Move over Merge will cause it to do so (assuming a version of Move over Merge which is a purely local principle, not requiring multiple convergent derivations in its domain).

(66)  a.  \[ [\text{TP someone [T to } [\text{VP be someone here}]]] \]
    b.  \[ [\text{TP T [VP seems [TP someone [T to } [\text{VP be someone here}]]]]] \]

(67)  a.  \[ [\text{TP someone [T T [VP seems [TP someone [T to } [\text{VP be someone here}]]]]]] \]
    b.  \[ N = \{\text{there}_1\} \]
    c.  ‘*Someone seems to be here’ + N = \{\text{there}_1\}’

(68)  a.  \[ [\text{TP there [T T [VP seems [TP someone [T to } [\text{VP be someone here}]]]]]] \]
    b.  ‘*There seems some one to be here.’

\(^{24}\) Shima makes no remark on whether his proposed Move over Merge principle, or his interpretation of MOM, operate only on sets of convergent derivations; therefore, we are left to assume, from his very strict focus on local decision-points, that he endorses a local version of the principles.
When we reach the point of filling the matrix [Spec, TP] in (66b), there are still two options for filling it: inserting *there*, which is still in the numeration, or moving *someone* up yet again (which we are allowed to do based on assumption (61a) above). If *someone* moves, *there* will be left in the numeration, which will thus be un-exhausted, and the derivation will crash ([67] above). If *there* merges instead, we end up with the ungrammatical sentence from (6b), repeated in (68b). According to Shima, “the preference for Move over Merge requires Spec of the matrix TP to be filled by raising [of *someone*] rather than insertion” (p. 383). He provides no explicit discussion of the case where *there* merges instead, but by the above quotation, presumably this derivational path crashes because it violates the Move over Merge principle.

In Shima’s explanation of these cases, it is important to note that assumption (61a) does a lot of the work. It would be fair to say that Shima’s account of the data is more a result of the particular Case assumptions he makes in (61a) – (61c) above than it is a result of an economy principle. Let us imagine what would happen, keeping all of Shima’s assumptions intact except the economy principle Move over Merge. First, taking the case where partitive case is assigned to *someone*, we have two options from the stage in (64). If we Merge, *there* is inserted, further structure is added, and *there* eventually moves up to the [Spec, TP] of the matrix clause, deriving *There seems to be someone here*. If we Move at (64) instead, *someone* moves to Spec of the embedded TP, thereby violating (61a). Perhaps the derivation crashes at this point, but if it doesn’t, there are other ways in which it can also crash. If *someone* then moves to the [matrix Spec, TP], then *there* will not be inserted, and will crash for having *there* remain in the numeration. If *there* is inserted into [matrix Spec, TP] then the derivation will violate the assumption in (61a) twice. Thus, under the assumption that *someone* receives partitive case, one derivational path independently leads to convergence, and one leads to a crash (i.e. no economy principle is motivated).

Under the second option, where *someone* is not assigned partitive case, we have the following possible derivations. First, regardless of whether we Move or Merge, both insertion of *there* at (64) and raising of *someone* should be compatible with assumption (61a). However, no derivation will be able to converge if *someone* does not have partitive case, since assumption (61c) stipulates the expletive needs to have the opportunity to check its posited formal feature with a DP with partitive case. Thus, if we take this principle seriously, all derivations under the option where *someone* does not receive partitive case will crash. Note that this crashing is again independent of economy. It thus appears possible to derive the contrast between (6a) and (6b) without economy at all, if we take all of Shima’s assumptions in (61) seriously.

How happy should we be with the stipulations Shima makes in (61)? They at least deserve further examination. The main criticism I make against them at this point is that they are stipulative, and little or no independent evidence or theoretical motivation is given for them. Assumption (61a) in particular seems very suspect. In (66), *someone* is allowed to move to check the EPP feature of the infinitival T0 in the embedded clause, by virtue of its having structural case, but it does not actually check case in this position. (If it did, then Shima’s analysis would violate a generally accepted idea that infinitival Spec TP’s do not check case.) It seems odd that the stipulation that allows it to merge in the embedded [Spec TP] is a property that will not be realized (i.e. valued as nominative) until it moves to the [Spec TP] of the matrix clause. Does *someone* enter the derivation in some way marked for having structural case? Furthermore, we should examine again the original proposal by Belletti (1988) and Lasnik (1995a,b) that copular *be* assigns partitive case, since there may well be empirical evidence against this proposal not discussed in Shima (2000).

---

25 E. Cowper (p.c.) notes that structural case is used in this derivation almost like a ‘diacritic’; I would add that, because of this, the solution Shima provides lacks explanatory power.
Shima (2000: 384, fn. 8) also includes comments by an anonymous reviewer that point out another shortcoming of his analysis: it cannot explain the distinction between the following sentences:

(69)  
\begin{itemize}
  \item a. It seems that there is someone in the room.
  \item b. *There seems that it is someone in the room.
\end{itemize}

The conceptual apparatus supplied by Shima’s analysis does not account for this contrast since “the preference for Move over Merge [proposed by Shima] has nothing to do with the choice between it-insertion and there-insertion.” Furthermore, one cannot argue that the ungrammaticality of (69b) is a result of it blocking the necessary association between There and its associate. This is, according to the reviewer, because it does not have partitive case. The two sentences in (69) have identical numerations, with the main distinction being that in (69a) there is the first expletive merged, while in (69b) it is the first expletive merged (to fill [Spec TP] of the embedded clause). Shima’s (2000: 384, fn. 8) tentative explanation is that “[Spec, T0], which selects the copular be with partitive Case, must be filled by there rather than it,” but obviously this begs the question of why.

Of the three assumptions (in [61]) that Shima makes, I find the idea in (61c) that the expletive must check a formal feature with an associate to be convincing, though mainly for reasons that will be discussed in section 3.4.3. However, the qualification that this DP must have partitive case still strikes me as stipulative and deserving of further investigation. Although Shima presents an alternative account of the core data in (6), he does so with the unnecessary crutch of a new economy principle (unnecessary because we can derive the same results without it, keeping his other assumptions constant, as discussed in the above paragraphs), and under some stipulative assumptions about Case that need further empirical support. Thus, we will turn elsewhere for new alternative analyses of the data in (6) that do not appeal to MOM.

3.4.2 Castillo, Drury & Grohmann

Castillo, Drury & Grohmann (1999) take up the issue of the original data in (6), and add other problematic cases, such as the following:

(70)  
\begin{itemize}
  \item a. *John seems that it was in the room.
  \item b. It seems that John was in the room.
\end{itemize}

(71)  
\begin{itemize}
  \item a. There was a [rumour [that a man was in the room]] in the air.
  \item b. [A rumour [that there was a man in the room]] was in the air.\footnote{Originally attributed to Juan Romero and Alec Marantz in Uriagereka (1999b), though according to the authors (Castillo et al. 1999), also discussed in Frampton & Guttman (1998).}
\end{itemize}

However, the crucial difference between the data in (6) (what Castillo et al 1999 call the ‘core case’) and the data in (70) – (71) above is that the data in (70) – (71) can be accounted for using other mechanisms, mechanisms which Castillo et al (1999) suggest have ‘independent plausibility’. These include Exhaust Array, phases, and kernel numerations (i.e. of Uriagereka 1999a). Castillo et al (1999) state that, “once the additional mechanisms are in place [e.g. kernel numerations, the operation Exhaust Array, etc.], the cases that seem difficult to initially account for (e.g. the Romero/Marantz examples) come out right even if there is no cost distinction between operations. This reinforces the interest of examining alternatives [to the cost-of-operations account for (6)]” (p. 21). Furthermore, though the examples in (70) – (71) above can be accounted for by “(cyclic)
constraints on derivations/lexicon access,” the original motivating data for MOM, (6), still remains unexplained. It is precisely because this contrast is the one specific, yet resilient, case that justifies MOM, that we are motivated to try to account for it without MOM. As Castillo et al (1999: 20) state: “[the contrast in (6)] seems (again) to stick out as the only empirical motivation for the cost-distinctions.” Castillo et al (1999: 20) reach the same conclusion as I, that cost-of-operation economy is suspect, though with different weight assigned to specific arguments: “the additional mechanisms (phases, etc.) discussed […], to the extent that they have independent plausibility, suggest to us that we should dispense with the cost-of-operations distinctions.” Their proposed tack, identical in broad strokes to what I have argued for above, is “[to deny] that both [6a] and [6b] are both well-formed (thus nullifying any potential comparison between them and thus removing the motivation for comparison).” I would then add that dispensing with recourse to MOM in this “core” empirical case argues very directly against MOM in general, which furthermore is a large piece of the picture for ‘economy principles’ generally, and certainly for cost-of-operation-based principles. With this in mind, let’s examine the details of Castillo et al’s (1999) Anti-EPP story.

If we get rid of the EPP, the analysis of (6) is almost over before it begins. Without EPP features, (6b) is straightforwardly ungrammatical. There is no reason, and therefore no justification, to compare the derivations of these sentences for economy purposes, since (6b) does not converge. Specifically, supposing an infinitival T0 does not “host a specifier,” there is no way to derive (6b), since we will not reach a decision point for the embedded clause where one can either insert the expletive or move the associate to fill this spot. However, we will still require the [Spec,TP] of the matrix clause to be filled since the T0 head will have uninterpretable ϕ features which need to be matched. T0 of the embedded clause, on the other hand, is defective and lacks ϕ with which to match against a nominal or DP. The only reason to fill this position is the EPP; so without the EPP, the decision point where Move or Merge is applied disappears.

What of the other decision point, though? The one then created by delaying the choice between expletive and associate, to the point of filling the [Spec, T0] of the matrix clause? In essence, one of these options won’t converge. Without the EPP, the derivation now looks like this:

(72) a. [TP to be someone here]
   b. N0 = {there1, T1, seem1, to0, be0, someone0, here0}

(73) a. [VP seems [TP to be someone here]]
   b. Nn+1 = {there1, T1, seem0, to0, be0, someone0, here0}

(74) a. [TP T0 [VP seems [TP to be someone here]]]
   b. Nn+2 = {there1, T0, seem0, to0, be0, someone0, here0}

At this point, we are met with the same decision that MOM determined earlier: whether to move the associate, as in (75a) below, or insert an expletive, as in (75b) below.

(75) a. [TP someone [T T0 [VP seems [TP to be someone here]]]]
   b. [TP there [T T0 [VP seems [TP to be someone here]]]]

However, the decision point only appears to be similar to our earlier conundrum. In this case, assuming there to be an expletive in the numeration, the two phrase markers directly above correspond to the following numerations:

(76) a. Nn = {there1, T0, seem0, to0, be0, someone0, here0}
   b. N0 = {there0, T0, seem0, to0, be0, someone0, here0}
The remaining *there* in the numeration for phrase marker (75a) indicates that this derivation will not converge; therefore, the two derivations are not comparable for the ‘global’ version of the MOM, and this false ‘decision point’ is really no such thing at all.

Castillo et al tackle this slightly differently, suggesting copying directly from the lexicon rather than selecting from a numeration (in line with Collins 1997), but the basic result is the same as that I have outlined above: “if there is any independent need for numerations/arrays or sub-divisions of these in terms of relativized convergence domains […] then the much more general condition ‘Exhaust Array’ will suffice to introduce the expletive if it is present in the set of initial lexical choices” (Castillo et al 1999: 21 – 22). If an expletive is present in the ‘set of initial lexical choices,’ but not introduced in the course of the derivation, then the derivation will crash. There is no need here to appeal to an economy condition like MOM.

A detailed discussion of all the repercussions of doing away with the EPP is beyond the scope of this paper, but the general opinion I adopt here is that while Castillo et al (1999: 21) may succeed in demonstrating that “the consequences of eliminating the EPP are surprisingly far from disastrous,” they have not succeeded in demonstrating that the payoff is worth it. Despite the ‘mysterious’ nature of the EPP as an explanatory apparatus that lacks explicit or transparent motivation, it is nonetheless worth exploring other non-economy stories for the core empirical case in (6). A preferable non-economy account of this data should also not do away with a core component of current theory, the EPP.

Furthermore, if their story holds, it is not clear how the existence of expletives can be explained at all. Expletives seem, from a descriptive standpoint, only to exist to fill the [Spec, TP] position of clauses. Castillo et al’s perspective is that expletives exist to ‘mediate’ the transfer of features between TI and the associate; in other words, “the intuition is that *there* serves to ‘transmit' both Case to the associate and the ϕ-features of the associate to T/INFL” (p.22). However, I think their discussion of this matter addresses the question of what the expletive is doing, when it is there, rather than the question of *why* the expletive is there at all. We turn next to a few recent accounts that share a derivational property: that the expletive originates low in the embedded clause, rather than being inserted in [Spec, TP].

### 3.4.3 Low-Origin Accounts: Hornstein (2009)/Witkos (2004) and Deal (2009)

There is considerable evidence for the idea that during the derivation of expletive sentences, the associate and the expletive at some point form a complex constituent. Milsark (1974), Chomsky (1986), Den Dikken (1995), Witkos (2004), Hornstein and Witkos (2003), and Hornstein (2009) have all observed properties that seem to point towards a close relationship between expletive and associate. Some of these properties are described below.

Witkos (2004), Hornstein and Witkos (2003) note that expletives and associates have properties that are “chain-like,” similar to properties observed in A-chains between antecedents and traces. For example, in (77) below the relationship between *there* and *someone/a beer* respectively in (77a) and (77b) is subject to the same kind of locality constraints as between *someone* and *a beer* and their traces in (78a) and (78b).

(77)  
   a. *There seems that someone is in the room.*  
   b. *There is the man drinking a beer.*

(78)  
   a. *Someone seems that t is here.*  
   b. *A beer is the man drinking t.*

27 Data in (77) – (78), (79a – c), and (80) from Hornstein (2009) and Witkos (2004).
The authors also note (as Milsark 1974 and others have before them) that there is a one-to-one correlation between expletives and associates, as in (79) below. (79c) shows you cannot have two expletive there’s with only one associate (the ‘too many there’s problem’ subsequently taken up by Deal 2009).

(79)  a. It/*There was preferred for there to be someone at home.
  b. It/*There was difficult for Bill for there to be someone at home.
  c. *There seems there to be someone in the room.
  d. There/*It was time for it to be possible to finish.

There is also a definiteness effect, as in (80), whereby the associate of there must be indefinite. In (80a) everyone behaves like a definite, for reasons related to inclusiveness (see Lyons 1999: 32).

(80)  a. *There is everyone in the room.
  b. *There is the man drinking a lot of beer.

In addition to these properties, Deal (2009) points out properties of expletive there based on types of verbs that can select it. She notes that this is not a much-discussed property of expletive there, having been catalogued but not explained in Levin (1993). (For previous discussion of these facts, though, one can refer to Deal’s extensive reference list on p. 2, fn. 2.) Deal argues that “there-insertion does not target predicates at random” (p. 2). So, for example, there can appear in tandem with the verbs listed in (81), but not with those in (82). This kind of selectional restriction on the predicates that there can combine with is evidence of a close relationship between the verb and the expletive, arguably one that is realized structurally.

(81)  a. There appeared a shadowy figure in the doorway.
  b. There arrived a train in the station.
     likewise: accumulate, coexist, emerge, hover, live, lurk, predominate, sit, swing

(82)  a. *There laughed a man in the hallway.
  b. *There melted a block of ice in the front yard.
  c. *There slowed a train on the eastbound track.
     likewise: abate, break, collect, detonate, divide, level, reddensify, vaporize

Deal (2009: 5) uses the above classificatory distinction to argue that “it is the functional structure surrounding verb roots which controls the insertion of there,” an analysis which entails the expletive originates low, rather than high, in the tree.

Hornstein and Witkos also expound the benefits of a low origin account. We can see from the data in (77) – (79) that such empirical effects provide evidence for the associate and the expletive forming a complex. It is more standard to assume that the expletive is merged high, and that the associate moves to join it (for example, at LF). But recent proposals by Hornstein (2009), Hornstein and Witkos (2003), Witkos (2004), and Deal (2009) argue for the low-insertion hypothesis.28 I will not present the arguments against high-origin expletive analyses that raise the associate. For discussion, see Hornstein (2009: 136 - 137), Witkos (2004: 177 - 180), and Deal (2009: 3). I will assume the low-origin account of expletives to be correct, and will be concerned here with describing and elaborating on the two recent proposals by Hornstein (2009)/Witkos (2004) and Deal (2009).

28 Other authors who have championed the low-insertion of English expletives are Groat (1999), Moro (1997), Den Dikken (1995), Hoekstra and Mulder (1990), and Zwart (1992). (List partially excerpted from Groat 1999.)
Enforcing the low-insertion hypothesis leads to a solution to the core data in (6) without appealing to Merge over Move.

Hornstein (2009) presents an account of the data in (6) that does not appeal to economy conditions; his solution goes a considerable way towards explaining the above-noted qualities of expletive constructions, although it also dispenses with (long-distance) AGREE. Hornstein (2009: 139) writes, “the A-chain properties of there/associate pairs ([77] and [78] above) follow straightforwardly if [([77])] is a case of A-movement by there… The one-to-one correlation between expletives and associates also follows on the assumption that there and the associate start off as a doubled constituent.” Thus, the expletive originates as in (83) – (84) below.

(83) [There is [[there someone] in the room]]. (Hornstein 2009: 139 / Witkos 2004: 181)
(84) [TP There [seems [TP there [to be [[there + someone] here ]]]]] (Witkos 2004: 192)

Under this analysis, there is “D-like,” taking an NP complement (in this case, someone). So there originates where someone originates, then raises. Note that under this analysis, we never reach the crucial “decision point” for MOM, repeated in (85) below. By the time we get to the stage in (85), both there and someone have already merged. There is no Merge versus Move.

(85) [TP T [seems [TP to be [[there +someone] here]]]]

However, though the mechanisms introduced above will allow us to derive (6a) (i.e. as in [84]), they do not solve the problem of how to rule out (6b). Witkos (2004: 191-192) notes two possible solutions:

“One is to reject the assumption that infinitives have Spec TP positions for elements to move into. If so, there is only one potential landing site for there […] the matrix [Spec, T]. […] The second possibility is that movement of the expletive without the whole associate is more economical because less is moved and so is preferred if convergence is so possible.”

This second suggestion is known in the literature as the Lighter Load principle, which Hornstein (2009: 40) says is suggested in Chomsky (1995) though not formalized there. This is another economy condition; since the exercise of these sections has been to discover whether economy principles are necessary to solving the problem of the data in (6), this is obviously not a solution we want to endorse without significant supporting evidence. The other solution Witkos (2004) suggests is Castillo et al.’s (1999) idea of dispensing with the EPP for infinitival clauses (also suggested by Epstein & Seely 1999). It would also be enough to simply stipulate that the associate does not move; but what we want to know is why it does not move, so this is not a very satisfying answer. Before we comment on these alternatives further, let’s briefly review the proposal by Deal (2009) and see where that leaves us.

Deal (2009) suggests, based on the patterns observed in (81) and (82), that verbal structure dictates whether there can be inserted (and it is always inserted low). Her causative hypothesis, in (86) below, along with further elaboration on the verbal structures surround verb roots (see Deal 2009: 13 – 14) suggests that there insertion is only possible in the specifier of non-inchoative unaccusatives.

(86) Causative Hypothesis: The vP of an unaccusative verbal root may contain expletive there just in case it does not contain CAUSE.
I will not review all of the details of Deal’s (2009) analysis. The upshot of it, for the issues in this paper, is that it provides evidence that expletives must be merged within the verbal domain, encouraging a low-insertion analysis, like Witkos (2004)/Hornstein (2009), and ruling out the possibility that there originates in the [Spec, TP] of the embedded clause in (6). Unlike Witkos (2004)/Hornstein (2009), Deal’s (2009) proposal would place expletive there in the specifier of a ‘default’ verbalizer (identified by the notational v~), a species of little v; Witkos and Hornstein would place there in a complex constituent with the bare NP that is its associate. Deal (2009: 17) also argues that “there has uninterpretable features which it checks against its associate,” and states that this is accomplished via a local Agree relationship. Though the two proposals differ in specifics, they both provide important evidence for the idea that expletives begin their derivational journey very low in the tree, and this essentially takes away the context in the embedded [Spec, TP] for the MOM story.

What remains to be explained, however, is why the sentence in (6b) is ill-formed. We have already seen Witkos’ suggestions above. Deal (2009: 21) suggests that what makes (6b), repeated below in (87), ill-formed is the associate’s movement to [Spec, TP] of the embedded clause.

(87) *There seems someone to be here.

Deal (2009: 22) suggests that the movement of the associate, someone, to the embedded [Spec, TP] might be ‘spurious,’ and that this position is ‘semantically lethal.’ She speculates that the “Spec,TP position of non-control infinitives is a ‘lethal A-position’, semantically speaking.” The data in (88) below is independent evidence that this position is ‘semantically lethal,’ an observation made by Moulton (2007).

(88) a. Stolen documents were alleged to be in the drawer.
   b. *John alleged stolen documents to be in the drawer.
   c. John alleged there to be stolen documents in the drawer.

I think further research into this possibility would potentially be very fruitful. In any case, we have succeeded in demonstrating that low origin accounts have significant empirical support, while also removing the original context—the decision point for Merge or Move in the embedded [Spec, TP]—for MOM for the core data in (6).

3.5 Discussion

There are several alternative stories to choose from if we want to eschew MOM and still derive the empirical results of the core case(s). As we have seen, there are various strategies, and I suggest that the most promising are those presented by Hornstein (2009)/Witkos (2004), and Deal (2009), which are not identical but also not entirely incompatible. It would certainly appear from this cornucopia of analyses that MOM is not necessary to derive the facts of the core case. Whether the alternative analyses presented here are convincing or not, and which is actually correct, may depend on empirical or theoretical arguments that have not yet been constructed, or on facts not yet discovered. But given that empirical evidence for MOM is scarce, and that the core case used in support of it can be solved in many other plausible ways, it seems worth spending some effort developing the consequences of a theory that does without it. The next section (Section 4) will address some residual empirical cases that have provided tenuous empirical support for MOM, and provide novel discussion on how the facts may follow without MOM, further weakening the already negligible empirical basis for the economy principle.
4. Residual Empirical Arguments for MOM

Though the core case of the data in (6) is by far the primary empirical evidence for the economy principle Merge over Move, there are a few other instances where the principle has been used to some empirical effect; I will refer to these cases as the ‘residual’ empirical arguments for Merge over Move. According to Hornstein (2009 and p.c. February 2009), there are two other residual cases that might, if the arguments and data are accurate, empirically justify MOM. These are (1) the argument for subject control into adjuncts, barring object control and assuming sideways movement (Hornstein 2001), and (2) the derivation of [object, subject] word order in Icelandic object shift (Jonas 1996, Chomsky 1998/2000). Neither of these arguments is particularly popular, although for different reasons. The first is sometimes discounted for its reliance on sideward movement, and in the second the veracity of the data itself is questionable (Hornstein, p.c. June 2009).

The following sections provide critical discussion of these residual cases, with the purpose again of assessing the strength of the empirical evidence for MOM, while also examining the consequences of reanalyzing them without using MOM. As in the previous section, I will demonstrate that MOM appears to be an unnecessary theoretical postulate in the few tenuous empirical cases where it can be argued to have an effect.

4.1 The empirical argument from subject control into adjuncts, assuming sideways movement

This argument for MOM comes from Hornstein’s (2001) proposed explanation of why it is not possible to get obligatory object control into adjuncts. For example, in (80) below, PRO can be co-indexed with John, but not with Mary.

(89) John saw Mary without PRO

The analysis provided in Hornstein (2001) could be called unorthodox for its use of sideways movement; for this reason, it is generally not considered as strong empirical evidence for the economy principle MOM. While controversial perhaps, sideways movement does not, I think, deserve to be a black mark for this case of MOM. In this section I assume that sideways movement has its merits, and thus present an account for the lack of object control into adjuncts using sideways movement but without with Merge over Move.

Hornstein (2001) proposes that this sentence is built up using several different ‘workspaces’: that is, separate subtrees are constructed in parallel, and the subtrees are then joined to construct the whole tree. Another key assumption of Hornstein’s program is that obligatory control is actually movement. We will adopt this assumption for the present. Thus, Hornstein proposes that the ungrammatical interpretation of (89) above is a result of the illicit movement of Mary to the complement position of the matrix verb, represented in (90b) below. Subject control into adjunct phrases is derived through movement of John to [Spec, V0], as indicated in (90a).

(90) a. John saw Mary without John entering the room.
b. *John saw Mary without Mary entering the room.

To summarize, the derivation of (90a) is achieved through multiple workspaces in the following way. First, the ‘adjunct’ is constructed; second, the matrix clause construction is begun. John then moves
‘sideways’ from the constructed ‘adjunct’ subtree into the ‘matrix’ subtree. The ‘adjunct’ is joined to the ‘matrix’ subtree so that the two subtrees are unified into one tree, and construction of the single tree is completed by John moving up to [Spec, TP] of the matrix clause. This is outlined in more detail below:

(91) ‘Adjunct Workspace’: (‘+’ indicates a Merge of two elements)
   a. the + room
   b. entering + [the room]
   c. John + [entering [the room]]
   d. -ing (Infl) + [John [entering [the room]]]
   e. John moves from [Spec, VP] to [Spec, IP]
   f. ‘Adjunct’ merges with IP to create AdjunctP

At this stage, the derivation in the ‘Adjunct workspace’ looks like this:

(92) [AdjunctP without [IP John [V· -ing [VP John [V· entering [DP the room]]]]]]

In the ‘Matrix Workspace,’ the derivation proceeds as follows:

(93) ‘Matrix Workspace’:
   a. saw + Mary = [VP saw Mary]

John then moves from the subtree in the Adjunct Workspace to merge with [Spec, VP] of the ‘matrix’ subtree, checking the verb’s external θ-role. At this point, the ‘adjunct’ subtree properly becomes an adjunct by merging with the VP in the ‘Matrix workspace,’ resulting in the structure in (94) below:

(94) [<VP,VP> [VP John saw Mary] [AdjunctP without [IP John [V· -ing [VP John [V· entering [DP the room]]]]]]]

Economy plays a significant role in this derivation. Hornstein points to several particular steps of the derivation where economy is important. First, when John, within the ‘adjunct’ subtree, moves up within that subtree ([91e] above), MOM is violated. Within the current array, we still have Mary available. Choosing to move John rather than merge Mary at this point violates MOM. Hornstein (2001:48) notes this, and gives the following explanation. Abiding by MOM at this point would result in a crashed derivation; therefore, the grammar violates MOM. The source of the crash would be the unchecked Case on John: another economy principle, Minimality (‘Shortest Move’ or the ‘Minimal Link Condition’ in Hornstein 2001) specifies that the active DP closest to the target will raise, so if, at step (91e) above, Mary were to merge into the ‘adjunct subtree’ to fill the [Spec, IP], instead of John moving, then Mary’s position above John in the (sub)tree would block John from raising out of the ‘adjunct’ subtree and into the ‘matrix’ subtree.

Here I bring up my first point of contention: if ‘sideward’ movement of John is, as Hornstein (2001: 47) explicitly states, “not to a c-commanding position,” then what version of Minimality

29 As Hornstein points out, it is critical that this clause is, before it’s joined to the ‘matrix’ clause, the ‘adjunct’ clause only by virtue of what it will become after it joins the other subtree. In order to avoid violating the Extension Condition, it is necessary that the ‘adjunct’ sub-tree is not actually an adjunct until it merges with the other (‘matrix’) sub-tree.
would restrict John’s movement, as opposed to Mary’s, once both DPs were in the ‘adjunct’ subtree? Both at this point have unchecked case features:

(95) \[
\text{[AdjunctP without [IP Mary\textsubscript{q}, CASE: ? [\textsc{v'}-ing [VP John\textsubscript{q}, CASE: ? [v entering [DP the room]]]]]]}
\]

Since the target is not in a c-commanding position, and both DPs in the ‘adjunct’ subtree are equally active, wouldn’t this make them equidistant from the target, or even make Minimality irrelevant for the present case? If this is the case, it is hard to follow Hornstein’s argument that MOM can be violated because Minimality would otherwise prevent convergence. It is not clear what definition of Minimality is assumed to operate here. In fact, Hornstein must make an argument for the irrelevance of the MLC/Minimality when Mary has been merged as complement of saw and John moves up to check the verb’s external theta-role. In this case, John and Mary are argued to be equidistant from the target (Spec VP of saw) because “‘Mary’ in the complement position of ‘saw’ and ‘John’ in the SpecIP of the adjunct do not c-command one another nor does the target of movement c-command them both” (p. 48-49). So while at this step, John and Mary are argued to be equidistant, there is an implicit assumption that they are not equidistant from the target if they have both been merged into the adjunct. It must be the fact that Mary c-commands John if merged into Spec IP of the adjunct clause that allows Minimality to have an effect, despite the fact that neither is c-commanded by the target/probe in the matrix clause.  

However, though this conclusion is convenient, it does bear further scrutiny. Furthermore, we should look warily on an analysis that requires us to postulate that some economy principles (e.g. MOM) are violable if their application would otherwise lead to a crashed derivation, while other economy principles are seemingly inviolable in such a situation (e.g. Minimality). What is it then, that makes Minimality a ‘stronger’ (i.e. less violable) economy condition than MOM? Why is it that, since these two ‘economy’ principles come into conflict in this derivation, that MOM must yield to Minimality? This is what the empirical effects seem to demand, yet what is the theoretical motivation for creating tiers of economy principles, some more violable than others?  

There also seems to be a kind of circularity to Hornstein’s argument for why MOM can be violated in the construction of the ‘adjunct’ subtree. The argument, again, is that John moves up to [Spec, IP] within the adjunct subtree, instead of merging Mary into this position, a violation of MOM. This violation is acceptable because abiding by the economy principle and merging Mary instead of moving John would lead to a crashed derivation later on, since Minimality would not allow John to move out of its merge position to check case. Thus, violations of MOM are acceptable if abiding by the condition would lead to non-convergence. I think there is a kind of logical circularity to this argument that is worth noting. Instead of acknowledging that the crashed derivation is empirical evidence against MOM, the reasoning here is that the grammar is allowed to violate the condition because the derivation crashes. Essentially, if MOM is enforced and the derivation crashes—a bad thing for MOM, one would think—this is actually okay for the economy principle precisely because the derivation crashes. Though it is conceivable that MOM is an economy principle with look-ahead properties such that it is violable if the derivation crashes, such a claim must be motivated beyond this particular case of adjunct control, since it is naturally suspect for its inherent circularity.

---

30 Hornstein (2001: 58, fn. 57) notes that “closeness” can be defined in many ways, and that sideways movement is more compatible with a definition of “closeness” for the purposes of Minimality that does not define “closeness” in terms of the target (probe).  
31 Particularly if economy conditions are meant to be specific instantiations of a more general, unifying concept of economy within the grammar, this idea of stratifying economy conditions seems implausible (as pointed out by Denis, p.c.).
Furthermore, we would want to demonstrate that the formulation of MOM here is falsifiable, in keeping with the principles of sound science. The MOM principle argued for by Chomsky (1995) is falsified in a case where, of two apparently convergent (and comparable, i.e. starting from the same numerations) derivations, one of which continues with Move and one of which continues with Merge, the Move derivation is the only grammatical output. The principle is also falsified under the condition that we have two convergent (comparable) derivations, one of which chooses Merge and another of which chooses Move at a given point, both of which have outputs that are grammatical. This is what we would expect if no preference based on cost of operation were active in the grammar. The principle, as formulated by Chomsky, is then a falsifiable claim. On the other hand, it appears that MOM as formulated in Hornstein (2001), as a violable principle that is, by rule, ignored if enforcing it will cause a crash, is not a falsifiable principle. Specifically, adding the violability condition to the principle makes it unfalsifiable. This is true because we cannot imagine a situation where the claim about MOM’s ‘defeasibility’ could be disproven. If the principle is violable where crashes occur, because crashes occur, what empirical situation would provide evidence against this claim?

The defeasibility part of this principle has the logical structure in (96). If, in two comparable derivations with the same numerations, at a given step S, there is the option of either merging or moving, and the merge path leads to non-convergence, then move. However, for this to be falsified, we need an empirical situation that has the structure as represented in (96b).

\[(96)\]
\[
\begin{align*}
\text{a. } & (\text{Merge path } \rightarrow \text{non-convergence}) \rightarrow \text{Move} \\
\text{b. } & (\text{Merge path } \rightarrow \text{non-convergence}) \rightarrow \text{Merge}
\end{align*}
\]

However, the grammar cannot perform the arrangement described in (96b), where the merge path, which leads to non-convergence, is ‘chosen,’ or else it can perform it trivially only insofar as it will produce an ungrammatical utterance. The grammar cannot provide a validation of the scenario outlined in (96b), since the grammar can only validate a path by giving a grammatical output. So the defeasibility of the MOM condition turns out not to make the condition unfalsifiable.

Note, further, that MOM in this case (i.e. its violation via the raising of John in the adjunct), does not range over convergent derivations, as in the original formulation of the principle in Chomsky (1995). It is not the case that we have two convergent derivations—one where Mary merges into the ‘adjunct’ subtree and one where John moves up to [Spec, IP]—between which MOM chooses the ‘most economical.’ In fact, what we have is one derivational path where a crash results (if we merge Mary) for independent reasons (i.e.: for example, if Mary is merged into the adjunct, one could argue that John will not be able to check case). In Hornstein’s (2001) argument for MOM based on subject control into adjuncts (which will be outlined in more detail below), the reasoning is more parallel to Chomsky’s (1995) analysis: in this case, we have two potentially convergent derivations. There is no obvious reason, accepting Hornstein’s assumptions, that the derivation where John moves sideways into the matrix subtree to be the complement of saw, should not converge. Thus there is no obvious reason why we should not have object control into adjuncts, given the assumptions. But if Hornstein wants to rid himself of this troubling case in the construction of the adjunct where Merge over Move is apparently violated, it might be enough for his purposes to say that Merge over Move is not relevant here because it does not range over convergent derivations. In this case, it is not that MOM is violable at this point, but that it is simply not relevant.

\[32\] Cf. Popper (1963: 36): “The criterion of the scientific status of a theory is its falsifiability.”

\[33\] As Denis (p.c.) points out, it is also possible that if a preference for Move or Merge exists in the grammar, that it is parameterized instead of being a core principle of Universal Grammar.
Back to our main empirical problem: why don’t we get object control into adjuncts? Why, when we have obligatory control (OC) into adjuncts, is it necessarily subject control? Take, for example, the data in (89) – (90) above, repeated here below:

(97)  
(a) John$_i$ saw Mary$_j$ without PRO$_e$$_j$ entering the room.  
(b) John saw Mary without John$_e$ entering the room.  
(c) *John saw Mary without Mary$_e$ entering the room.

The adjunct, without entering the room, can only be interpreted as applying to John, the subject of the matrix clause, and not to Mary, the object. The reason for this, according to Hornstein (2001), is the direct result of Merge over Move. At stage (92) above, where the ‘adjunct’ subtree is complete, and the ‘matrix’ subtree is about to undergo its first merge, Mary is still available in the numeration. It is therefore more economical, according to the MOM principle, to merge Mary from the numeration than to move John sideways into the ‘matrix’ subtree. The specific step which is thus blocked by MOM, where John moves out of the adjunct subtree to merge as complement to the verb in the matrix subtree, is represented in (98) below:

(98)  
(a) [AdjunctP without [IP John [I'-ing [VP John [V' entering [DP the room]]]]]]$_e$adjunct  
(b) [VP saw John]$_e$matrix  
(c) N$_e$(b) = {Mary, v, I$_e$}

Assuming $N_0$ (the initial numeration) includes both Mary and John, John will never merge as the complement of saw in the ‘matrix’ subtree. It will only merge as the external argument of saw, with the result that obligatory control into the adjunct clause will be from the subject of the matrix clause, and never the object. That both Mary and John are in $N_0$ is the assumption I will tinker with in my own analysis below. Thus, for Hornstein, MOM is the reason we don’t get object control into adjuncts. However, I don’t think this is convincing enough evidence for MOM, even if we do accept sideways movement (and obligatory control as movement) in this case. I think there is a more convincing argument to be made that other independently necessary considerations are enough to rule out the possibility of John merging as an object in the matrix subtree.

Hornstein uses MOM to explain lack of object control into adjuncts; below I present an argument that this empirical effect can be derived in one of two ways (one of which I will argue is preferable), neither of which requires the use of the economy principle.

The crucial step in the derivation that MOM rules out, thereby ruling out object control, is the movement of John out of the ‘adjunct’ clause to the complement of saw, while the movement of John to the external argument position of the matrix verb is acceptable. This follows if Mary is merged first with saw, before John moves. There are two other straightforward ways of accomplishing this, rather than using MOM to stipulate that the merge of Mary is more economical than the move of John. The first is to stipulate that the ‘matrix’ subtree must be constructed first. This in fact follows from other, more reasonable and less stipulative assumptions about the construction of these subtrees. If we assume that the subtrees are constructed in ‘strict parallel,’ meaning that it is possible to at least roughly measure out ‘steps’ to derivations, and that each ‘step’ (i.e. each merge or move or application of some other operation) is roughly equivalent, lengthwise, then we end up at the natural endpoint of having [VP saw Mary] be constructed before the ‘adjunct’ subtree is even halfway finished. Thus, by the time John is ready to move sideways out of the ‘adjunct’ subtree into the ‘matrix’ subtree, the only position available to it will be in [Spec, VP], as the external argument to saw:

34 Contrary examples will present themselves shortly.
The second strategy, which is more defensible than the first, and which in fact might be implicit in the first, is that separate workspaces must work from separate subarrays. This seems very intuitive, and for that reason one might be suspicious of it, but there does seem to be some theoretical motivation for postulating that if we allow separate workspaces, then we must also allow for partitioning of the numeration.

### 4.1.1 Problems with building subtrees from a non-partitioned numeration

Several formal problems are raised by assuming that when separate subtrees are being constructed, it is necessary to work from the same numeration, without any partitioning of that numeration. For one thing, if, as Hornstein (2001: 75) suggests, the ability for Merge to operate on multiple subtrees in parallel is independently necessary in the theory as it is, then having separate subtrees select from separate numerations would greatly reduce the space of possible derivations that the grammar needs to navigate through. Assume, as Hornstein does, that the derivation outlined in (91) – (94) above selects from one and the same numeration. Assume also that Merge is blind: it does not ‘know,’ before concatenating two elements, whether this concatenation is viable or not (i.e. whether it will lead to a crash or not). Note that this assumption is independent of whether or not the grammar has lookahead properties. The grammar might be able to ‘cursorily’ sketch a set of derivations for comparative purposes—for example, to determine whether a certain path crashes or not—but the operation Merge does not itself contain a lookahead ability. So even though the grammar might have mechanisms by which it can acquire knowledge about certain derivational paths versus others, the operation Merge concatenates elements ‘blindly,’ (i.e. without such omniscience). Merge does not know whether the two components it merges can be merged, though the grammar may have mechanisms by which it determines, over sets of derivational paths, which are going to converge or not. So, there are no restrictions on Merge.

Having separate subarrays for separate subtrees of the derivation would reduce the navigational space of the grammar by reducing the combinatorial possibilities made available by Merge. Given a smaller subset of numeration elements to select from, and the assumptions outlined above, the number of crashed derivations would be reduced by the use of sub-arrays. The benefit of this rests on an implicit assumption that crashed derivations do in fact use up resources. If we assume a grammar with lookahead capabilities, it is assumed that the calculation of combinatorial possibilities, and their outcomes, also uses up resources. It would therefore be in the spirit of Minimalism to adopt the use of subarrays, since overall this reduces the resources used by the grammar in computing derivations.

If separate arrays were an entirely new concept, adopted only in these cases of sideward movement, this might put their adoption here in tension with methodological economy, which legislates against the unnecessary addition of principles or theoretical apparatus; however, subarrays seem to be empirically required in cases that do not involve sideways movement. Specifically, subarrays have been proposed, most famously by Chomsky (1995), to explain how expletives can end up in the matrix clause in sentences with embedded complement clauses. Such a situation should

---

35 I elaborate further on the second strategy. The first seems less defensible for several reasons, including the fact that it is not clear how the grammar is to know that the matrix tree is constructed first, or how this is to be carried out in a derivation, without inventing some kind of communication mechanism between two workspaces (E. Cowper p.c. points this out). The idea of counting steps “in strict parallel” likewise does not seem particularly feasible, in part because it introduces more questions than it resolves: for example, how and why this parallel timing is to be enforced, and how the grammar measures a ‘step.’
not arise if MOM is assumed. However, without assuming MOM, subarrays are not well-motivated empirically by expletive data. We will return to this in the final section of this paper.

4.2 Deriving subject control into adjuncts without MOM, assuming a partitioned numeration

Armed with the possibility of separate sub-arrays for separate subtrees, and with the independently necessary condition, Exhaust Array (see Castillo et al 1999, Uriagereka 1999, etc.), we can see how interarboreal movement between subtrees might be restricted to occurring only after the subarrays for individual subtrees have been exhausted.\(^{36}\) The derivation of (90a) would involve the following subarrays.

\[(100)\]

\[
\begin{align*}
\text{a. Numeration: } & \{\text{John, Mary, saw, entering, the, room, without, T, v, -ing}\} \\
\text{b. } & \text{Array}_{\text{matrix}}: \{\text{Mary, saw, v}\} \\
\text{c. } & \text{Array}_{\text{adjunct}}: \{\text{John, entering, without, the, room, -ing}\} \\
\text{d. } & \text{Array}_{\text{composite}}: \{T\}
\end{align*}
\]

Array\text{composite} is the array accessed after the completed ‘adjunct’ subtree merges with the ‘matrix’ subtree which has exhausted all elements in its subarray. It seems necessary to posit that the ‘matrix’ subtree array does not contain all of the (in this case, functional) elements that will be necessary to complete the tree once the ‘adjunct’ subtree merges with the ‘matrix’ subtree. If the T of the matrix clause were included in the (sub)array\text{matrix}, then the subarray would necessarily not be exhausted before sideways movement took place, and it would not be possible to use Exhaust Array to rule out John’s ever merging interarborally into the complement of the matrix verb.

Below, I outline how the derivation would proceed under these assumptions.

The ‘matrix’ subtree and the ‘adjunct’ subtree begin independent construction in separate ‘workspaces’, or separate ‘islands’ within one large workspace, working from separate subarrays as defined in (100) above. Assume the two subtrees are constructed independently and in parallel.\(^{37}\) The ‘matrix’ subtree is constructed by merging Mary and saw, then ‘light’ v. This results in the structure in (101), and the Array\text{matrix} is exhausted.

\[(101)\]

\[
\begin{align*}
\text{a. } & \left[\text{VP } v [\text{VP saw [Mary]]}\right] \\
\text{b. } & \text{Array}_{\text{matrix}} = \{v_0, \text{Mary}_0, \text{saw}_0\}
\end{align*}
\]

The ‘adjunct’ subtree is constructed from the Array\text{adjunct}, until it is exhausted, as in (102).

\[(102)\]

\[
\begin{align*}
\text{a. } & \left[\text{AdjunctP without } [\text{TP John [T: ing [VP John [v: entering [DP the room]]]]]}\right] \\
\text{b. } & \text{Array}_{\text{adjunct}} = \{\text{without}_0, \text{John}_0, \text{entering}_0, \text{the}_0, \text{room}_0, \text{-ing}_0\}
\end{align*}
\]

Note that at no point is Mary ever available to merge into this adjunct subtree. This removes the problem that Hornstein was forced to solve by positing that MOM played a role at this stage. Now since both arrays have been exhausted, the grammar is left with the choice of selecting a new array from the original numeration, or selecting a component from the completed ‘adjunct’ subtree for sideways movement. It could, also, I suppose, at this point select the entire adjunct for merger.

\(^{36}\) Though the proposal suggested below is a plausible explanation, it is in need of further empirical evidence to be soundly justified.

\(^{37}\) The other option is that operations like Merge operate on the workspace as a whole, and thus two things cannot be merged at the same ‘time,’ despite their belonging to separate subtrees.
leave the details of these mechanisms to be worked out at a later date. Sideways movement of John out of the ‘adjunct’ is now possible since both subarrays have been exhausted, and movement of John satisfies the external θ-role of v in the matrix subtree, and also allows John to eventually check case in [Spec, TP] of the matrix clause. So the movement is well motivated (at least in sense of “enlightened self-interest”). If John does not move sideways, the matrix verb saw will not be able to discharge its external θ-role, and John will not be able to check its case.

After John has moved sideways to [Spec vP] of the matrix subtree, resulting in the structure in (103), the ‘adjunct’ subtree is then selected by Merge operating on the matrix subtree, resulting in (104).

(103) \([_vP \text{John} [\_v \cdot v [\_vP \text{ saw Mary}]]]\)

(104) \([_{<_vP, vP>} [_{>_P, vP} [_{>_P, vP} [_{>_P, vP} [_{>_P, vP} [_{>_P, vP} [\_v \cdot v [\_vP \text{ saw Mary}]]] [\_\text{AdjunctP without [TP \text{John} [\_v \cdot \text{ing} [\_vP \text{John} [\_v \cdot \text{entering} [\_vP \text{the room}])))]]]]]])]

At this point, nothing more is available to be operated on by Merge (or by Move); thus, a new array must be made available to the derivation. This array is as in (100d), containing the single functional head, T (or I). The result is the completion, and convergence, of the derivation, with John raising to [Spec, T] to check T’s ϕ and EPP features, as in (105).

(105) \([_{TP \text{John} [T \cdot T [_{<_vP, vP>} [_{>_P, vP} [_{>_P, vP} [_{>_P, vP} [_{>_P, vP} [_{>_P, vP} [\_v \cdot v [\_vP \text{ saw Mary}]]] [\_\text{AdjunctP without [TP \text{John} [T \cdot \text{ing} [\_vP \text{John} [\_v \cdot \text{entering} [\_vP \text{the room}])))]]]]]]]])

Using subarrays which separately feed the independent subtrees in this derivation has the same result as Hornstein’s use of MOM and a single numeration: object control into the adjunct is ruled out in this case. In Hornstein’s analysis, this was the result of a preference condition for merging Mary as the complement of saw rather than moving John (sideways). In the analysis presented above, this falls out from the fact that Mary is a part of a subarray for the matrix subtree which must be exhausted prior to sideways movement. John and Mary start in separate arrays; by the time John can merge into the matrix subtree, the complement position of saw is already taken.

What would happen, though, if Mary and John weren’t nicely partitioned into separate arrays as they are in (100)? It might be possible to find independent evidence for the idea that separate subtrees must be built from separate lexical arrays, but, faced with the current absence of such evidence, let’s look closely at what would happen in the derivation in a few other possible cases.

Firstly, if Mary appears in the Array_adjunct in (100) with John, one of two things can happen. Keeping our other assumptions stable (i.e. there is no Merge over Move), Mary can either be merged into [Spec TP] of the adjunct clause, or John can be raised to this position, leaving Mary in the array. If John moves up, then Mary will remain in the adjunct array and the unexhausted array will cause

38 Further research is needed to determine what triggers the grammar to select an entire adjunct or an element from another sub-tree, or to activate a new sub-array. The discussion here indicates where sideways movement is possible, but not under what conditions it will always occur (i.e. what triggers it). It could be that a new array is not activated at this point precisely because John is available for sideways movement, since it remains caseless.

39 Note that allowing John to move into a theta-position in the matrix subtree violates standard assumptions about the illegality of movement to a theta-position. However, Hornstein assumes that theta-roles are similar to other kinds of features the grammar checks, and thus allows John to move to check a theta-role/feature. For the present, I will assume this to be correct.

40 This is slightly different from Hornstein’s (2001) original analysis, where the adjunct was attached at the VP level (2001: 48).
the derivation to crash. (Or, alternatively, if Mary never merges into the derivation, one of the theta roles of the matrix verb, see, will remain undischarged: John can move sideways to receive one of these theta-roles but not both.\(^{41}\)) On the other hand, if Mary merges into [Spec, TP]\textsubscript{adjunct} and John fails to move up to this position, this will interfere with John’s checking case (i.e. it will be unable to do so), and the derivation will crash at a later stage. There are also theoretical problems with merging Mary into the [Spec, TP]\textsubscript{adjunct} position: this is not a theta position, and even though I am willing to accept, in the derivation outlined in (100) – (105), that John might acquire a theta role during the course of the derivation, I am not similarly willing to accept that original merge of an element into the tree can happen in non-theta positions (e.g. Spec, TP). In (100) – (105), John is merged into a theta-position and then acquires an additional theta-role from the matrix verb, saw. However, merge of Mary into the adjunct clause at [Spec, TP] does not result in Mary acquiring a theta-role at initial merge.

What if Mary were in the Array\textsubscript{composite} instead of the Array\textsubscript{matrix} in (100)? This is potentially a more serious complication for the derivation that has been proposed above. In this case, we would predict that John, from the adjunct, would become the internal argument of saw, since Mary would not be available to fill this position. This makes the wrong prediction, incorrectly deriving object control into the adjunct. However, one advantage of relegating Mary to the matrix array (and thus ensuring it becomes the object of saw) is that then all of the (matrix) verb’s theta roles can be discharged prior to activating the composite array. This means that within the matrix array, the components necessary for that phase (the one delimited by v) to be thematically complete are all available. Conceptually, it seems preferable that Mary, as an argument, should appear within the same array as the verb from which it gets its theta role upon first merge. It thus seems conceptually unwarranted to partition Mary into a higher array (i.e. the Array\textsubscript{composite}).

Putting this derivation aside, there do appear to be instances of object control into adjuncts that are not yet accounted for, exemplified in (106) below.

(106) a. John punished Bill for driving the car.
    b. John arrested Bill for driving drunk.

In Hornstein’s (2001) analysis, these exceptions are noted and cursorily dealt with, but an extensive analysis is not undertaken. Hornstein speculates that in the cases above, the adjunct is attached lower than in the derivations sketched earlier. We could take this to be a strike against the MOM account of subject control into adjuncts, since the MOM account Hornstein (2001) gives would also seemingly rule out sentences like (106). Furthermore, it is not transparent why having the adjunct attached lower in the tree would clear up this case for the Merge over Move story. I leave these cases to further research.


There is a further case, (also noted by Hornstein 2009: 50) where MOM has been argued to have some empirical bite. The original Icelandic data is found in Jonas (1996), and presented as ‘virtual Icelandic’ in Chomsky (1998/2000). Hornstein (p.c.) notes that the data that the argument is based on is quite possibly incorrect, and has therefore been considered controversial. I will not give it an in-depth treatment here, but present the basic argument from Chomsky (1998/2000).

In the derivation of the virtual Icelandic sentence in (107), the derivation presumably reaches the point represented in (108a), where little v has been merged above the verb phrase and its direct

\(^{41}\) Although, having an unexhausted Array\textsubscript{adjunct} may interfere with the triggering of sideways movement, preventing John from moving out of the adjunct at all.
object, but the external argument has not yet been introduced. If, at this point, we assume that MOM has the power to ‘decide’ whether the external argument, *many students*, will merge, or whether the direct object will move, then we arrive at the correct order for Icelandic object shift: object precedes subject. Chomsky (1998/2000: 102) states that the feature that moves the DO to a Spec of little $v$ is an EPP-feature by analogy to this feature of $T$. Assuming an “at-point” preference for Merge over Move, the external argument will merge before the internal argument moves, as in (108b) below. Then, the internal argument moves, (108c), $T$ merges, (108d), and the verb raises, (108e).

(107)  [The books]$_{DO}$ [many students]$_{SUBJ}$ read.

(108)  a.  $[v \ [\text{read} \ [\text{the books}]]]$

   b.  $[[\text{many students}]]_{SUBJ} \ v \ [\text{read} \ [\text{the books}]_{DO}]$

   c.  $[[\text{the books}]_{DO} \ [\text{many students}]]_{SUBJ} \ v \ [\text{read} \ t_{DO}]$

   d.  $[T \ [\text{the books}]_{DO} \ [\text{many students}]]_{SUBJ} \ v \ [\text{read} \ t_{DO}]$

   e.  $[\text{read} \ [T \ [\text{the books}]_{DO} \ [\text{many students}]]_{SUBJ} \ v \ [\text{read} \ t_{DO}]$

The question is whether the derivation that chooses Move instead of Merge at point (108a) above actually converges or not. If it does, then we might have a resilient empirical case similar to (6), where of two apparently convergent derivations, only one is actually grammatical. If it does not, then MOM is not necessary to account for the facts described above.

4.4 Summing up the Residual Cases

Above, I have summarized some residual empirical cases which have been argued to provide empirical support for Merge over Move beyond the traditional empirical argument from the data in (6). For the case of Icelandic object shift, further work is necessary to show conclusively whether MOM is necessary or not to account for the data, though Hornstein (p.c. June 2009) notes that “the Icelandic data [which is used to support Merge over Move] is considered very controversial and likely wrong.” As for the argument from subject/*object control into adjuncts, there is evidence against the idea that Merge over Move prohibits control of objects into adjuncts from the data in (106); it may be possible to account for these cases and leave the MOM story intact for the other data (i.e. sentences like those in [89 – 90]). However, this must be argued for; it is not obvious. Beyond this analysis-internal evidence against MOM, there is also evidence, as with the famous expletive data, that MOM is simply not needed for the data in (89 – 90) to derive the proper effects. This further stacks up the evidence against MOM—that the principle is either wrong, or it is epiphenomenal to the mechanisms of the grammar, or it is simply unnecessary, and therefore ruled out by the principles of methodological economy.

5. Theoretical Consequences and Conclusion

This paper has presented empirical and conceptual arguments against the economy principle, Merge over Move, with critical discussion of economy principles more generally and our responsibility to hold them to stringent empirical standards. Particularly, this paper has argued that MOM has primarily one case of empirical justification, and that this one case has several possible alternative explanations which do not appeal to economy. Furthermore, I have argued that in the few rare empirical cases, besides (6), where MOM has been argued for, it is possible to account for the data without Merge over Move. Throughout, I have argued for shifting the burden of proof to demonstrating that Merge over Move is in fact an empirically necessary component of UG, and away from proceeding with analyses on the basis of assuming it.
What is particularly troubling about the unsteady empirical (and conceptual) grounds for Merge over Move is that this economy principle, and the case in (6) that motivates it, has been called empirical motivation for significant theoretical expansions of apparatus in syntax: specifically, subarrays and phases. I think I have provided enough motivation here to put Merge over Move to rest in favour of rethinking the original cases that were its empirical justification. What becomes of subarrays and phases is then quite a serious matter, and one which bears much more scrutiny, beyond the scope of this paper. Is there empirical motivation for subarrays? There seems to be some evidence for phases and a related idea, multiple spell-out, (though Boeckx and Grohmann 2007 have pointed to the inadequacies of current definitions of phases), but what about the ‘input’ end of the derivation? What evidence, once Merge over Move is put to rest, is left for subarrays?

As an illustration of the potential theoretical consequences of removing MOM from our model of UG, let’s consider the examples in (109) – (111). Chomsky (1998/2000: 104) has proposed that sentences like the one in (109) – (111) are the empirical basis for postulating the existence of subarrays (and phases).

(109) It was decided PRO to be executed $t$ at dawn.\footnote{Note Chomsky’s (1998/2000: 105) remark that the grammatical interpretation of this sentence is that “the prisoners decided that they would be executed at dawn, but not that we decided that they would be.”} (data from Chomsky 1998/2000)

(110) There was assumed to be a reason why a man is in the garden. (data from Anagnostopoulou and Fox 2007)

(111) There was a rumour that a man, was $t_i$ in the room. (Boeckx and Grohmann 2007:208)

The argument is as follows: if Merge preempts Move, then we would expect never to be able to derive sentences like (109) - (111), where an expletive must be in the initial numeration, but not merge in the most embedded clause. In (109), the expletive should merge to satisfy the T of the embedded clause [to be executed at dawn], instead of PRO moving. In (110), the expletive should merge to satisfy the T of the embedded clause [is in the garden], rather than a man moving. Likewise, in (111), merger of there should preempt movement of a man to fill the [Spec TP] position of [was in the room]. If Merge over Move is to be obeyed, then this necessitates the introduction of subarrays to partition the expletive into the higher clause—in other words, to make the expletive unavailable to merge into the most embedded clause. Chomsky also uses this as the motivation for phases, but Boeckx and Grohmann (2007: 208) point out that what this really provides evidence for is lexical arrays (subarrays).

Interestingly, if we ignore MOM, we are able to derive sentences like (111) without difficulty, if we merely allow the grammar to have the option of moving or merging at the key point of filling [Spec, TP]. More specifically, if at the point of filling [Spec, TP] of [was in the room] in (111), the grammar can either insert an expletive or move a man, then it is conceivable that we end up with two options, represented in (112) below, both of which are convergent to this point, and have the possibility of converging as full sentences down the line, if further clauses were then constructed above these (e.g. in [113]).

(112) a. [there T [was a man in the room]]
   b. [a man, T [was $t_i$ in the room]]
(113)  a. [TP John said [CP that [TP there was a man in the room]]]

These facts give plausibility to the idea that without Merge over Move, the need for subarrays and the numeration should be reexamined.

Given the weakness of the case for Merge over Move, at least in any of its current instantiations, we should reexamine the independent evidence for the numeration and for (sub)arrays. Since MOM is a key piece of evidence for such theoretical apparatus, this paper motivates a return to searching for further evidence of the necessity of subarrays and the numeration. In doing this, it is important to divorce the notion of subarrays and the numeration from the notion of spellout by phase, or multiple spellout (or some kind of highly cyclical spellout). The distinction is between the input end of the derivation and the output end. Some kind of phase-based approach in the output (i.e. spellout) seems to be empirically well motivated (I leave examples to further research); separation by subarray on the input, end, however, is less well motivated, and therefore we should place renewed effort into discovering solid empirical foundation for such a concept. One place to begin looking for independent evidence for subarrays might be in empirical analyses that make use of sideways movement, and the idea of separate workspaces for complex specifiers. A partitioned numeration for separate derivational workspaces was suggested in Section 4 as one way to derive desired empirical results.

I suggest that we have, at this point, substantial empirical and conceptual evidence for laying MOM to rest, and meanwhile casting a wary eye at other proposed economy conditions, being vigilant of the same weaknesses we have discovered for MOM: lack of consistency with ‘global’ versus ‘local’ instantiations, defining the economic ‘metric’ in an ad hoc way, and a lack of comprehensive empirical support.

Chomsky (1998/2000: 100, 101, 103, inter alia) repeatedly states that these very theoretical issues are, at their heart, empirical. Here I have tried to give a broad and refined sense of what the empirical state of the evidence for MOM is. The outlook for MOM does not look promising, at least in any of the forms that it has hitherto been proposed. Perhaps a novel conception of the idea will arise and resuscitate it with concrete empirical evidence. Perhaps, on the other hand, MOM is a distraction from the active mechanisms of the grammar that lie beneath it, giving the illusion of an ‘economy principle.’ It may even be that stronger empirical evidence exists for some version of a Merge over Move principle, or some kind of formulation of economy that is sensitive to the proposed simplicity of the operation Merge in comparison to other possible operations. However, it should then be a priority to find such evidence, since as I have demonstrated, this evidence is currently conspicuously lacking. If, as Chomsky (2000: 100) asserts, “the issues are empirical,” then it is time we converted this orthodox piece of wisdom into substantive results. Furthermore, it is always a beneficial theoretical exercise to examine basic assumptions of our model more closely; given a different constellation of assumptions, what will we find?

References


Uriagereka, Juan. 1999a. Minimalist Issues. Seminar, University of Maryland, College Park. Class lectures, papers, and other materials online available at: http://www.umd.edu/Linguistics/People/Faculty/JUriagereka/home.html


