8

Constraining the Learning Path without Constraints, or The OCP and NOBANANA[1]

CHARLES REISS

That which is wanting cannot be numbered.

Ecclesiastes 1: 15

[S]o far as I can tell the story is always more or less the same: whenever there is behavior of significant complexity its most plausible explanation tends to be some explicit process of evolution, not the implicit satisfaction of constraints.

Stephen Wolfram, *A New Kind of Science* (2002: 351)

8.1 Introduction

Many linguists, especially phonologists, have assumed that both Universal Grammar and particular grammars contain constraints, *qua* prohibitions on grammatical structures.[2] However, such prohibitions *cannot* be learned by positive evidence (an infinite number of well-formed structures are absent from the PLD—we may find a supposed ill-formed structure in the next

[1] Thanks to audiences at McGill, UQAM, MIT, Utrecht, and Michigan and to Andrea Gormley, David Odden, Jean-Philippe Marcotte, Daniela Isac, Geoffrey Pullum, Marshall Wong, Brendan Gillon, Patrick Davidson, Bill Idsardi, Madelyn Kissock, Jonathan Bobaljik, Glyne Piggott, Detmar Meurers, Eric Raimy, Ida Toivonen, Ash Asudeh, Morris Halle, and Bert Vaux for useful comments and criticism, not all of which have been addressed. Some of these people remain outraged by the contents of this paper. Many of these ideas were developed in conversations with Mark Hale. It should be clear that many of my proposals draw on various ideas in the literature, some that precede and some that overlap with the seven-year gestation period of this paper. I do not pretend to have approached exhaustive acknowledgement of these sources.

[2] Some developments in Minimalist syntax are discussed below.

sentence we encounter). Therefore, these prohibitions could only be learned *via* negative evidence.

However, it is generally accepted that negative evidence is neither supplied to the child with sufficient regularity, nor attended to by the child enough when supplied, to play a significant role in language learning. Therefore, since the prohibitions cannot be learned *via* positive evidence (for reasons of logic), nor through negative evidence (according to the empirical data), they must be innate.

This conclusion follows from the premises, but I believe it is false. The fault lies with the assumption that UG, and also particular grammars, consist of constraints.

In this chapter, I justify rejection of this premise, and I demonstrate how the need for constraints can be circumvented, while still allowing a learner to converge on a grammar in a finite amount of time. In other words, the learning path is constrained, but not because of language-specific or universal constraints. We thus escape from the tendency to develop overly rich models of Universal Grammar that have culminated in recent theories such as Optimality Theory.

This chapter thus has two goals. One goal is to argue that well-formedness constraints are inappropriate computational devices for modeling grammar. Thus the chapter attempts to do in phonology what recent work by scholars such as Samuel Epstein (Epstein *et al.* 1998; Epstein and Seely forthcoming) is attempting in syntax—to develop a purely derivational theory with minimal theoretical apparatus and no filters or well-formedness constraints. Similar ideas are discussed by Szabolcsi (1988). The conceptual arguments will be bolstered by reference to recent work developing alternative approaches to phonological computation from constraint-based ones.

The second goal is to make concrete proposals concerning the nature of phonological acquisition. The idea is to constrain the acquisition task without recourse to innate constraints.

8.2 The universal NoBanana constraint

Let's turn to a preposterous example. Suppose we are seeking a constrained theory of UG for syntax and we are trying to choose between a theory with the components in (1a) and another with the components in (1b):[3]

[3] I am obviously making simplifying assumptions here. The point is just that one model has a set of entities and the second has all those plus an additional constraint.

(1) Which model of UG is better?

a. *Merge*
Lexicon

b. *Merge*
Lexicon
NoBanana

Model (a) contains the rule *Merge* which operates on elements of the *Lexicon*. Model (b) contains these components as well as the additional constraint NoBanana which marks as ungrammatical any representation of a sentence containing a banana—an actual banana, not the lexical item *banana*. Is it useful to claim that (b) is a more constrained model than (a) is, since (a) has no way of ruling out sentences that contain bananas? Obviously it is not useful or necessary to do this—(a) does not generate sentences that contain bananas since bananas are not contained in the set of items (the *Lexicon*) over which *Merge* operates. The more constrained model is thus (a), since it is characterized by a subset of the elements needed to characterize (b), and the two models have the same extension.

Consider another preposterous example in (2).

(2) Which model of UG is better?

a. *Merge*
Lexicon

b. *Merge*
Lexicon
NoLexicalItemsSentToCleveland

In (1) we considered the effect of enriching a model of grammar by adding a constraint referring to entities not found in the set over which *Merge* applies. In (2b), we have added a constraint referring to an operation that is not present in the model of the grammar in (2a). Again, it should be clear that since *Merge* does not have the effect of sending lexical items to Cleveland, and since the grammars characterized in (2) contain no other operations, it is not necessary to rule out representations in which lexical items have been sent to Cleveland.

What makes the preceding examples preposterous is that constraints are supposed to be formulated in terms of a (typically implicit) *universe of discourse*. Note that the claim intended by the constraint NoBanana, that no representation of a sentence contains bananas, is probably true for all human languages. However, there are an infinite number of true claims of this type. No language requires speakers to dance a jig to express iterativity; no language has pizza as an element of syntactic trees; etc. Bananas, pizza, dancing of jigs, sending and Cleveland are not elements of grammatical models. In other words, we do not want our model of grammar to express every true statement about what structures do not occur, since there are an infinite number of such statements and the grammar must be statable in finite terms if it is to be instantiated in human brains.

The conclusion suggested by the preceding discussion is that the search for UG should be conceived of as the attempt to characterize the universe of discourse, the entities and operations that constitute the representations computed by the language faculty. UG is thus to be characterized by a list of categories and rules that take these categories as arguments—and nothing else.

A coherent conception of the "perfection" of the language faculty, one that does not cave in to the temptation of functionalism, is that the formal system that defines UG, as well as every particular grammar, is exhaustively definable: there is a finite list of categories and rules that uniquely determines all and only possible linguistic structures.[4] Again, UG should not be conceived of as a set of constraints defining directly what is *not* a possible human language, because this set has an infinite number of elements. The notion of what is not a possible language will follow from an appropriate characterization of the properties of possible languages, but this notion need not be independently formulated in the grammar.

8.3 Overview

This chapter not only develops this argument concerning what UG should not be, but also makes concrete suggestions concerning how the study of UG should be approached. In Section 8.4, I define constraints in opposition to rules, then I return to the issues raised in the Introduction in order to point out two slightly different ways in which inviolable constraints have been used. I then turn to a discussion of violable constraints, as used in Optimality Theory (Prince and Smolensky 1993). I conclude on philosophical grounds that linguistic theory should be rule-based rather than constraint-based: grammars contain rules (as defined below), not constraints (as defined below).

[4] In other words, the definition of UG, and of particular grammars, can be understood as including a final, exclusion clause of the type used in recursive definitions in logic. I address below the problem of overgeneration—the fact that the set of possible linguistics structures is a superset of attested structures.

In Section 8.5, I briefly show that the ideas presented here converge with some recent work in syntax. I then discuss, in Section 8.6, the use of constraints in conjunction with rule-based phonology, concentrating in Section 8.7 on the Obligatory Contour Principle (OCP) for illustration. Following Odden (1988) I argue that there is no good theoretical or empirical motivation for positing the OCP. The argument extends readily to other constraints that have been posited in the literature.

Section 8.8 compares rule- and constraint-based approaches to phonology. I argue for a revival of rule-based phonology, but not a return to the mixing of rules and constraints, and I offer a contribution to the understanding of *formal* aspects of Universal Grammar. The results presented here demonstrate that progress in our understanding of UG does not depend upon the characterization of substantive tendencies subsumed under the notion of markedness. Some ideas concerning the acquisition of phonology, and how learning can be constrained without constraints are presented in section 8.9. Conclusions and open questions are discussed in section 8.10.

8.4 On constraints

This section discusses in general terms various uses of the notion of constraint in linguistic theory. First I discuss the distinction between rules and constraints. Then I discuss constraints on grammars, that is, constraints on what is a possible language. I then turn to inviolable constraints within grammars. Next, I discuss violable constraints as the basis of grammatical computation, as in Optimality Theory. I argue that each of these approaches to defining UG suffers from a combination of a lack of elegance and a mistreatment of the problem of inductive uncertainty.

8.4.1 *What is a rule? What is a constraint?*

Mohanan (2000: 146) argues that, due to basic logical equivalences, the constraint/rule distinction is incoherent once we adopt the view that both rules and constraints express propositions. However, in the following definitions I distinguish rules and constraints both in terms of their role in a computational system (a grammar) as a whole and in terms of their putative "grounding" in phonetics.

8.4.1.1 *A system-internal definition of rules vs. constraints* Various practices in the literature may be at odds with the definitions developed here. This purely terminological issue does not bear on the validity of the dichotomy proposed. So, for example, we may find formal statements that are called "constraints" in the context of a given theoretical framework, but which are in fact examples of what is here called a "rule." In some work, e.g., Karttunen (1993), the terms "(declarative) rules" and "constraints" are used interchangeably.

A RULE R can be viewed as a function that maps an input representation I defined in terms of a set of representational primitives (features and relations) to an output representation O which is defined in terms of the same set of primitives. The application of a rule depends upon a potential input representation matching the structural description of the rule. This representational matching procedure (RMP) outputs two possible results: YES, I satisfies the structural description of R; or NO, I does not satisfy the structural description of R. If the output of the RMP is YES, R applies and relevant parts of I are rewritten as O. If the output of the RMP is NO, I is not affected.

In a constraint-based theory, constraints also contain RMPs that serve to map an input I to one of the two possible results YES or NO, as above. However, for each constraint, one of the two values, YES or NO, maps to a further evaluation called VIOLATION and the other to NOVIOLATION. For example, in various versions of Optimality Theory, syllables that fulfill the condition expressed by "Does the syllable have a coda?" map to YES, and in the case of this constraint, YES maps to the value VIOLATION. Syllables without codas map to NO, which for this constraint maps to NOVIOLATION. For the constraint corresponding to the condition "Does the syllable have an onset?", a syllable with an onset maps to YES, which maps, for this constraint, to NOVIOLATION, whereas an onsetless syllable maps to a NO that, for this constraint, maps to VIOLATION. Perhaps this two-step evaluation of constraints is not necessary, but this characterization does reflect the widespread practice of using both negatively stated ("Don't have a coda") and positively stated ("Have an onset") constraints in the literature.

The use to which this evaluation as VIOLATION or NOVIOLATION is put rests with another part of the computational system. Violation of a constraint is passed on to other parts of the computational system. In theories incorporating inviolable constraints, constraint violation prevents a representation from being evaluated as grammatical. In Optimality Theory the violations are used by EVAL, the evaluation procedure which interprets violation with respect to the relative ranking of the constraints.

To reiterate: a rule is defined as a function from representations to representations; a constraint is defined as a function from representations to the set {VIOLATION, NOVIOLATION}.

The role of constraints in a computational system, telling another part of the system that a representation is somehow ill-formed, is related to the issue of NOBANANA, discussed above, as follows. There are an infinite number of ways in which a representation can be ill-formed. We do not want the grammar to have to be able to recognize them all.

8.4.1.2 *The system-external basis of well-formedness constraints* In many constraint-based linguistic theories a crucial aspect of constraint evaluation leading to the equivalent of an output value VIOLATION is the notion of ill-formedness or markedness. This represents the second major problem with constraint-based formalism, as defined here.

Depending on the formulation of a given constraint, either matching or failing to match the structural description of the constraint signals ill-formedness. To use the examples introduced above, a constraint formulated as "Don't have a coda" leads to an evaluation of ill-formedness for a syllable which *has* a coda, but a constraint formulated as "Have an onset" leads to an evaluation of ill-formedness for a syllable which *does not have* an onset. Relative and absolute ill-formedness or markedness evaluations of linguistic representations are ascribed by linguists for grammar-external reasons. Marked or ill-formed structures typically are claimed to have at least one of the following properties:

(3) Markedness criteria

- Relative rarity in the languages of the world
- Late "acquisition" by children (typically referring to the recognizability of a form in child speech)
- Loss in aphasia (typically referring to the recognizability of a form in aphasic speech)
- Relative difficulty of perception (not always experimentally validated)
- Relative difficulty of articulation (again, sometimes based on impressions of what is hard to say)
- Tendency to be lost in language change and to not arise in language change

All of these criteria have been criticized by Hale and Reiss (2000*a*, *b*; see references therein). These works conclude that the best way to gain an understanding of the computational system of phonology is to assume that the phonetic substance (say, the spectral properties of sound waves, or the physiology of articulation) that leads to the construction of phonological entities (say, feature matrices) *never* directly determines how the phonological entities are treated by the computational system. The computational system treats features as arbitrary symbols. What this means is that many of the so-called *phonological universals* (often discussed under the rubric of markedness) are in fact epiphenomena deriving from the interaction of extragrammatical factors like acoustic salience and the nature of language change. Phonology is not and should not be grounded in phonetics since the facts which phonetic grounding is meant to explain can be derived without reference to *phonology*. We return to these issues in Section 8.4.7 and in the critique of the Obligatory Contour Principle later in the chapter.

It should be noted that the papers by Hale and Reiss are by no means unique in arguing against the use of "substance" in determining the limits of grammar. A particularly clear example, pointed out to me by Ash Asudeh, is Kaplan (1995/1987: 346-7):

> A formal theory may have a relatively smooth outline ... [t]hen you start taking chunks out of it ... because you claim that no human language or grammar has such and such a property. ... It's a mistake to carry premature and unjustified substantive hypotheses into our computational and mathematical work, especially if it leads to mathematically complex, even if more restrictive, theories. ... [W]e should be wary of the seduction of substance.

The issue is related, as well, to discussion in other areas of cognitive science. For example, the almost universally held notion that segments that are (allegedly) complex from the articulatory perspective are also representationally complex or marked seems like a clear parallel to the error discussed by Pylyshyn (2003: 8) concerning work on vision: "the mistake of attributing to a mental representation the properties of what it represents." Again, we can relate our discussion to the issues raised by the proposed NoBanana constraint. Why do bananas make representations bad? Because they are not part of the system under scrutiny. But why should a grammatical constraint refer to something that is not part of grammar at all? More commonplace constraints, such as NoFrontRoundVowel, are as poorly motivated a part of Universal Grammar as NoBanana once we recognize that phonetic substance *cannot* be encoded in the phonology. The *acoustic* properties of front, rounded vowels are not directly accessible to the grammar.

8.4.2 *Karttunen (1993)*

I will not review all the literature debating the status of intermediate levels of representation, distinct from both input and output forms, that has appeared in the history of phonology, especially that focusing on Optimality Theory and its immediate predecessors and contemporaries. However, a few comments concerning the influential article of Karttunen (1993) are in order. I think that careful consideration will show that much of the debate concerning sequential vs. parallel derivation is empty.

Karttunen discusses the fact that phonological rewrite rules can be implemented by a finite state transducer. One advantage of expressing rules in this fashion is that transducers express relations between inputs and outputs bidirectionally, and thus can be more immediately useful in developing processing models for both production and parsing. A further advantage is that transducers corresponding to single rules can be composed into a single transducer that implements a "cascade" of ordered rules (p. 180). Thus, the ordered rule format and the transducer format are alternate means of expressing phonological knowledge. The intermediate representations of traditional rule-based phonology need not have a real-time processing referent—they can be understood as corresponding to the contribution made by each component transducer of a complex transducer.

It is interesting to examine Karttunen's ultimate explanation for a turn to two-level models with neither complex transducers nor ordered rules: "the composition of large rule systems to a single transducer turned out to be unfeasible because of practical limitations. A single transducer encoding the complexities of a language like Finnish was too large for the computers available in the early 1980s" (p. 180). Available computational resources have increased significantly over the last two decades, but in any event, such technological considerations are not obviously relevant to the evaluation of psychological theories, especially to the rejection of models that allow reference to intermediate levels of representation.

Whatever the status of the two-level models that arose from such considerations, there are a few points of interest in the context of this chapter. As Karttunen states, the "most fundamental aspect of the two-level rules is that they are deontic statements about correspondences that are possible, necessary, or prohibited in a certain environment"; they are "modal statements about how a form can, must or must not be realized" (Karttunen 1993: 181). In other words, the rules/constraints of the two-level models that Karttunen discusses are purely formal statements, not grounded in phonetic substance. Thus, the arguments used to motivate such a model do not necessarily extend to markedness-based models like Optimality Theory.

8.4.3 *Constraints on grammars*

It is a commonplace in the linguistic literature to find statements suggesting that a goal of linguistic research is to define UG by formulating the constraints on what is a possible language. This enterprise is typically seen as integral to explaining the paradox of language acquisition, in the following way. If the child is endowed with innate knowledge of the constraints delimiting the set of humanly attainable languages, then the child's hypothesis space is limited. Instead of choosing from the infinite set of (not even necessarily attainable) grammars, the learner need only select from a predetermined subset of those. Of course, we might make this idea more palatable to some by referring to constraints on the learner's ability to make hypotheses, rather than to knowledge of these constraints, but this is just a matter of terminology. I wish to argue that a characterization of UG in terms of such constraints can be at best merely a derivative notion.

It is necessary to stress that I am concerned in this subsection with constraints *on* grammars, not constraints *in* grammars. I am not concerned, for the moment, with evaluating the merits of constraint-based computational systems such as Optimality Theory (Prince and Smolensky 1993) *vis-à-vis* rule-based grammars, for example, although I turn to this topic below.

Instead of the preposterous examples in (1) and (2) above, consider the question of hierarchical structure in syntax. Let's imagine that we want to express the claim that all structure is hierarchically organized as a trait of UG. How should this proposal be formulated? If one seeks to characterize UG by listing constraints on the set of possible languages, then one might say something like "Flat structure is not possible" or "All structure is hierarchical". Again, since UG is instantiated in real brains, it must consist of a finite set of characteristic properties. Note again, that we would actually need an infinite set of constraining statements to characterize UG—those referring to bananas, jigs, etc. Again, there are an infinite number of such constraints on the set of possible languages. In order to avoid having an infinitely long list of constraints, constraint-based theories need a *list* of positive statements of entities (distinctive features, primitive operations like *Merge*, etc.). This list will define the universe of discourse in which we interpret a constraint like "Flat structure is not possible". We see, then, that a theory which formulates linguistic universals in terms of constraints must *also* contain a vocabulary of elements and operations in which those constraints are expressed, or to which they refer. This vocabulary of items and processes is presumably based on empirical observations and inferences. Consider a simpler alternative.

If our current hypothesis concerning UG is stated only in *positive* terms, as statements of what grammars have access to or consist of, without prohibitions or constraints, we can achieve a more economical model. The positive terms are just those entities and operations (features, deletions, insertions, *Merge*, *Move*, etc.) which have been observed empirically or inferred in the course of model construction. When faced with a phenomenon which is not immediately amenable to modeling using existing elements of the vocabulary, scientific methodology (basically Occam's Razor) guides us. We must first try to reduce the new phenomenon to a description in terms of the vocabulary we already have. If this can be shown to be impossible, only then can we justify expanding the vocabulary.

Thus, a "constraining approach" to UG, stated in terms of what is disallowed, requires a set of constraints, as well as a vocabulary which defines the universe of discourse in which the constraints are valid. The alternative proposed here requires only the vocabulary of possible entities and operations, along with the metatheoretic principle of Occam's Razor. The alternative is thus more elegant and should be preferred.

In more concrete terms this means that our theory of UG should consist of the minimum number of primitives that we need to describe the grammars we have seen.[5] Note that we should not be influenced in our search by preconceived notions of simplicity. For example, if we know that we need hierarchical structure for some phenomena, but there exist other phenomena which are ambiguous as to whether they require flat or hierarchical structure, then we should assume that the ambiguous cases also have hierarchical structure. If our current theory of UG contains an operation to generate hierarchical structure from primitive elements, constraints against flat structure will be superfluous. In fact, positive statements like "Structures are organized hierarchically" and "All branching is binary" (assuming they are correct) are also superfluous within the grammar itself, even though they are descriptively accurate, since they are just a reflection of how structure building operations work (see Section 8.5).

The approach advocated here seems to be consistent with that used in science in general. If a physicist observes a "constraint" on the behavior of a particle, say, then s/he posits a set of properties for that particle from which the observed behavior emerges. The constraint thus has the status of a derivative and not primitive aspect of the theory.

8.4.4 *Inviolable constraints in grammars*

It was suggested above that the issues raised thus far are irrelevant to the choice between rule-based and constraint-based computational systems. In a sense this was an overstatement and the discussion above is in fact clearly relevant to a certain class of constraints invoked in versions of Optimality Theory, as well as other models of phonology: constraints that are never violated,

[5] According to Rennison (2000: 138) this principle has, in practice, been more vigorously upheld by proponents of Government Phonology (GP), than by members of other schools of phonology.

either universally or within individual grammars. For the sake of concreteness let's adopt a version of Optimality Theory which assumes that it is never the case that the winning candidate in a derivation, in any language, has crossing association lines.[6] There are several ways to deal with this. One possibility is to claim that there exists a constraint, NoCross, that is part of the OT constraint hierarchy which incurs a mark when a candidate contains crossing association lines. This constraint can be posited to be universally undominated, or rather, universally undominated by a "competing" constraint. A competing constraint which dominated NoCross would be one whose satisfaction could "force" a violation of NoCross in the winning candidate. This possibility can be construed as allowing simplicity in the theory—allow GEN to generate candidates freely, and leave it to universally undominated constraints like NoCross to rule out candidates with no chance of surfacing. However, the simplicity achieved is somewhat illusory.

This approach introduces a complication into the core idea of Optimality Theory, the idea that grammars are defined by constraint hierarchies. If one adopts the view that constraints are universal and innate, then certain constraints, the undominatable ones like NoCross will have to be kept in a separate stratum of the constraint hierarchy, one whose members are not subject to reranking. Equivalently, they can be marked as not susceptible to reranking.

Yet another approach is to claim that these constraints are high-ranked at the initial state of the grammar. According to the claim of Smolensky (1996) and most other scholars, they would therefore start out at the top of the block of initially high-ranked Well-formedness constraints. If one is willing to accept such a scenario,[7] then the undominatable constraints need not be marked as unrerankable, since, by hypothesis, no language ever has evidence that they are dominated. However, the generalization that OT grammars consist of freely rerankable constraints becomes empty, if in fact, some of the constraints are never reranked in any language.

We see then that each of the versions of undominatable constraints proposed here leads to complications in the theory of grammar. An obvious

[6] This is a particularly well-known and easily discussed constraint. However, Local and Coleman (1994) have demonstrated that it is basically contentless.

[7] But see Hale and Reiss (1998) for arguments that it is untenable. They argue that acquisition under such an initial ranking, with Well-formedness constraints outranking Faithfulness constraints, is impossible. They claim that the (normal, rerankable) Well-formedness constraints must start out ranked below the Faithfulness constraints in order to allow the acquisition of a lexicon. If one adopts this assumption, then, the undominated Well-formedness constraints like NoCross would have to be initially ranked in a block separated from all the rerankable Well-formedness constraints, or somehow marked as not rerankable.

alternative is to state the constraints as limitations on Gen. In other words, assume that Gen freely generates—except that it does not generate forms that violate NoCross and other undominatable constraints. But this still fails to solve the need to define the universe of discourse for Gen. We would need constraints on Gen to keep it from generating representations that violate NoCross, but not ones that violate NoBanana, presumably. But Gen has certain properties, it does certain things with inputs, and we should try to characterize those properties. Therefore, it seems preferable to model Gen in such a way that it does not have the capacity to output forms with crossing association lines and other impossible traits (including bananas). In other words, the arguments against constraints on grammars and undominatable or inviolable constraints in grammars are the same—we always need a positive characterization of the formal system we are modeling.

8.4.5 *Free generation and constraints as filters*

The dominatable, or violable, constraints of both standard OT, which assumes universal, innate constraints, and other theories which allow language-specific constraints, do not immediately appear to pose the problems discussed thus far. Such constraints are formal devices for evaluating candidates, but they do not, each on its own, define what is a possible linguistic representation. However, I will argue in this subsection that even a constraint-based grammar which contains violable constraints is to be avoided. In section 8.7, we will see that the original motivation for such constraints may have been empirically and methodologically misguided.

Various theories of grammar, including Optimality Theory and some versions of Minimalism and its predecessors posit a mechanism that allows unconstrained generation of linguistic representations. In OT this device is Gen which, given an input, generates the universal candidate set of possible outputs. In various syntactic theories, an analog to Gen is the "free" concatenation of morphemes, or the "free" application of operations such as *Move α*. A derivation which is thus generated will either satisfy certain conditions at PF and LF, the grammar's interface levels, and thus *converge*; or it will not satisfy those conditions and it will *crash*. Both the OT approach and the free-generation-with-interface-conditions approach in syntax are flawed in the following (related) ways.

First, it is easy to proclaim something like "Gen generates any possible linguistic representation" or "The syntactic component allows *Move α* to apply freely." However, it is not clear what such statements mean. One could argue that the theory of grammar need not be computationally tractable, since grammar models knowledge and does not necessarily map directly to an algorithm for generating grammatical output. However, it does not follow from this that we should immediately aim for a model that we cannot imagine being implemented in the mind. It seems that any implementation of Gen or the syntactic component that incorporates *Move* α will have to be very explicit about what it does. One way to achieve this is to be explicit about what the abstract grammar generates. Second, the *free generation-cum-filters* model stinks somewhat of antimentalism. It basically says "We don't care how the candidate forms are generated, as long as they are generated. One way is as good as the next, as long as they are *extensionally* (empirically) equivalent." This is parallel to the position taken by Quine (1972, discussed by Chomsky 1986) in arguing that it is incoherent to talk about the "correct" grammar among a class of extensionally equivalent ones. In defining I-language, a matter of "individual psychology" as the domain of inquiry for linguistics, Chomsky (1986) argued convincingly that the fact that knowledge of language is instantiated in individual minds/brains means that there is necessarily a "correct" characterization of a speaker's grammar (or grammars). We will see below (Section 8.9.3) that such antimentalism does, in fact, show up in current theorizing. Once one accepts that modules/processes, like Gen and *Move* α, must have a certain set of properties; and that these properties ultimately must be derived from a set of positive statements (a vocabulary); and that these properties can be incorporated into the structural descriptions of rules; it appears to be the case that a procedural, or rule-based approach to grammar that generates a sequence of representations constituting a derivation is to be preferred to a constraint-based, non-derivational theory. In other words, grammars can be understood as complex functions mapping inputs to outputs. A rule-based model just breaks the complex function into simpler components, in order to understand the whole. A theory that incorporates Gen or *Move* α avoids the problem of characterizing the function that is the grammar.

8.4.6 *A New Kind of Linguistics?*

Many of the ideas presented here in support of rule-based grammars appear to be paralleled by claims made in Wolfram's (2002) *A New Kind of Science*. Wolfram has almost nothing to say about cognition, and he certainly did not intend his cellular automata models to apply to human grammars, but the logic is very similar to what I am aiming for and it is thus worthwhile quoting him fairly extensively.

In the following passages, equations can be understood as paralleling systems of constraints, and programs as paralleling systems of rules.

(4) Wolfram (2002: 368) arguing against constraints

> It is in many respects easier to work with programs than with equations. For once one has a program, one can always find out what its behavior will be just by running it. Yet with an equation one may need to do elaborate mathematical analysis in order to find out what behavior it can lead to. It does not help that models based on equations are often stated in a purely implicit form, so that rather than giving an actual procedure for determining how a system will behave–as a program does–they just give constraints on what the behavior must be, and provide no particular guidance about finding out what, if any, behavior will in fact satisfy these constraints.

Wolfram is suggesting that constraint-based analyses may provide a valid level of description, but leave unanswered certain important questions and also leave a certain amount of indeterminacy in understanding the nature of the system the constraints describe.

Basically the same point was made earlier in the book too, but from the perspective of the positive attributes of rule systems:

(5) Wolfram (2002: 342) on benefits of rule systems

> One feature of programs is that they immediately provide explicit rules that can be followed to determine how a system will behave. But in traditional science it is common to try to work instead with constraints that are merely supposed implicitly to force certain behavior to occur.
>
> ...I gave some examples of constraints, and I showed that constraints do exist that can force quite complex behavior to occur. But despite this, my strong suspicion is that of all the examples of complex behavior that we see in nature almost none can in the end best be explained in terms of constraints. The basic reason for this is that to work out what pattern of behavior will satisfy a given constraint usually seems far too difficult for it to be something that happens routinely in nature.
>
> Many types of constraints...have the property that given a specific pattern it is fairly easy to check whether the pattern satisfies the constraints. But the crucial point is that this fact by no means implies that it is necessarily easy to go from the constraints to find a pattern that satisfies them.
>
> The situation is quite different from what happens with explicit evolution rules. For if one knows such rules then these rules immediately yield a procedure for working out what behavior will occur. Yet if one only knows constraints then such constraints do not on their own immediately yield any specific procedure for working out what behavior will occur. In principle one could imagine

> looking at every possible pattern, and then picking out the ones that satisfy the constraints.

Given an input to a rule-based phonology, it is typically straightforward to compute the output (as long as the rules are explicit), whereas the problem of finding just which forms best satisfy a constraint system has proven difficult. These passages appear to be particularly relevant to discussions of the computational tractability of Optimality Theory, such as Idsardi (2006). My point in citing Wolfram is not to appeal to authority, but rather to show that the issues faced by linguists may be fruitfully compared to issues in other sciences where it is possible to simulate and model behavior with a greater degree of control.

8.4.7 *The fallacy of imperfection*

> It ain't why, why, why. It just is.
>
> Van Morrison

In phonology at least, it appears that the obstacle to developing such a theory has been an *a priori* belief in the relative well-formedness of abstract representations based on the never formalized notion of markedness. In other words, even the rule-based phonological literature is rife with constraints which are meant to "motivate" the application of rules that repair structure. In syntax, the tradition of appealing to markedness is more subtle, but it has basically been adapted in that the grammar, or perhaps the processor, is characterized with respect to derivations which "crash," as well as with respect to ones that "converge." Consider for comparison the visual system. Given an input, the visual system is assumed to have certain biases, probably manipulable via the little-understood mechanism of *attention*, but no visual input leads to a failure to assign a representation. It is also not clear what it would mean to say that a given representation generated by the visual system was less well-formed, or more marked than another representation. Presumably the visual system generates representations based on the input it is given, and these representations are unique—they are the best and the worst (or rather, neither best nor worst) that the system generates. Outputs are generated which depend on the input and the state of the system processing the inputs—hardly a controversial view. The same holds true of phonological representations—they are not perfect or imperfect, THEY JUST ARE. Since the violable OT constraints are posited on the basis of cross-linguistic typology, data from child speech and the informal intuition of linguists, it is worth evaluating these criteria. I do so here only briefly. Defining markedness based on cross-linguistic *tendencies* of absolute and implicational patterns of attestation (e.g., If a language has voiced stops, it also has voiceless ones) raises many difficult issues, not least of which is "How do we count?" Do we count tokens? E-languages like "English" or "Chinese"? Grammars?[8] Without an explicit theory of what gets counted, generalizations based on intuitive "statistical" patterns are worthless. Furthermore, at least some of the reported statistical tendencies, such as the more common absence of [p] from voiceless stop inventories, in comparison with [t] and [k], are highly reflective of areal biases in the sampling procedure (see Engstrand (1997) and Hale and Reiss (2000*a, b*) for discussion).

Hale and Reiss (1998) have argued in detail that the use of child speech data to determine markedness status is flawed since this data is rendered opaque by the effects of children's performance systems. I will not repeat these arguments here. Linguists' intuitions concerning "better" (unmarked) and "worse" (marked) structures reflect a confusion of levels of analysis, as well as other conceptual problems. A problem addressed in detail by Hale (2000) is that discussion of the evaluation of "output" forms often fails to distinguish between the output of the grammar (a feature-based representation) and, say, the output of the speaker (an acoustic or articulatory event). As demonstrated most clearly by our ability to construct 3D representations based on a black and white pattern on a printed page, there is a vast gap between physical stimuli and outputs and the representations that relate to them. Therefore, even if phonologists had a metric of the complexity or difficulty inherent in interpreting or creating certain physical stimuli or outputs (which they do not), it is apparent that there is no reason to believe that such a scale would translate straightforwardly to a markedness scale for representations. There is no reason to believe, for example, that the representation of the act of pushing a boulder is more difficult or complex or marked than the representation of the act of pushing a feather. Again, recall Pylyshyn's warning about "the mistake of attributing to a mental representation the properties of what it represents."

[8] I am collapsing Chomsky's discussion of a sociopolitical conception of "language," common in everyday parlance, with the E-language conception which he includes among the scientific approaches to the study of language. The E-language approach treats a language as an external artifact, say a text or corpus of texts, rather than as a knowledge state. This collapse is, I believe, justified and consistent with Chomsky's views, since the decision to include various texts or utterances within a single E-language corpus is typically made on the basis of the everyday sociopolitical notion of language—how else can an E-linguist decide that a set of texts constitutes a single corpus, except by appealing to the pretheoretical notion that they are all French or English or Swahili?

8.4.8 *OT constraints as fallible intuitions*

> We should know that one intrinsic characteristic of a heuristic is that it is *fallible*, and that it may be unjustified.
>
> Massimo Piatelli-Palmarini, *Inevitable Illusions* (1994: 22)

The preceding discussion suggests an explanation of why the constraints of OT are violable. These constraints are for the most part derived from so-called "principles of well-formedness" or "markedness" found in other phonological theories. I propose that these "principles" are actually just the heuristic devices that constitute our intuitions as experienced linguists. For example, we may assume that a sequence like [akra] will more likely have a syllable boundary before the stop-liquid cluster than between the two consonants. This is because we seem to believe, rightly or wrongly (it is hard to imagine how to collect the appropriate statistics under the I-language approach) that the majority of languages "maximize onsets" in such cases and leave the first syllable without a coda. However, both syllabifications are found, for example, in the Ancient Greek dialects. Lacking information to the contrary, it may be useful to assume that the more common syllabification is present in a new, unfamiliar language. This will allow the formulation of hypotheses that may then be tested, and the guess will turn out to be correct more often than not, if our intuitions have any basis. However, we must take care not to confuse our intuitions concerning what happens often with the actual nature of the system under study. Based on our experiences and expectations, we apply our intuitions in attempting to solve the problems involved with analyzing data, but there is no reason to expect that these intuitions directly reflect the nature of the actual mental grammar constructed by a learner. The intuition that heavy things fall faster than light things is very useful when someone drops something from a window, but the intuition needs to be transcended to understand the workings of gravity. Heuristics are used by the analyst to make useful guesses about data, and guesses can be wrong. This is why OT constraints need to be violable—they reflect the fallibility of our guesses.

It may be useful to refer to the error under discussion as a confusion of epistemological issues (concerning the nature of our knowledge) with ontological ones (concerning the nature of phonological systems). One explanation for the pervasiveness of such errors may lie with our terminology. A term like *physics* or *phonology* is used in a systematically ambiguous fashion. *Physics* means both "the study of the properties of the physical world, including gravitational attraction, etc." and "the properties of the physical world, including gravitational attraction, etc." When I fall down the stairs, I do so, not because there is a field of study that concerns itself with gravity, but because of the nature of the physical world, because of gravity itself. I would fall down the stairs even if all the physicists and physics books disappeared—I assume people fell down the stairs before Newton. By failing to make this crucial distinction we can be misled into believing that the *tools* (intuitions) we use in phonology *qua* field of study of the nature of sound systems are constitutive of phonology *qua* the nature of sound systems.

I think the use of violable wellformedness or markedness constraints in OT that are based upon putative statistical tendencies has exactly the status of this kind of error. Reiss (2000) discusses another such case in the OT literature.

8.4.9 *Overgeneration*

Pylyshyn (1984: 205ff) describes a box emitting certain recurrent patterns of signals. He then asks what we can conclude about the nature of the computational mechanism inside the box, based on the observed pattern of output. The answer is that we can conclude nothing, since the observed patterns may reflect the nature of what is being computed (in his example, the output is a Morse Code rendering of English text, and the observed regularity is the "i before e, except after c" rule), not the nature of the computer. In Pylyshyn's words "the observed constraint on [the system's] behavior is due not to its intrinsic capability but to what its states represent." (The observed "constraint" on output vanishes if we input German text, instead of English, since German texts will have words that violate the English spelling rule.) Pylyshyn's example suggests that we should expect our models to overgenerate with respect to the corpus of attested data, since this data is "sifted" by language change, for example. The language faculty may be able to perform computations that we have not observed because of the forms that language data just happen to take. The solution to this situation is clear: posit the minimal theoretical apparatus needed to generate attested patterns, and don't worry too much about overgeneration. We must assume, as a matter of scientific practice, that newly encountered phenomena will be amenable to modeling using current theories. We may be proven wrong, this is in the nature of inductive reasoning. When we are proven wrong we change the assumptions. This is not a bad situation—it just reflects the eternal incompleteness of scientific knowledge.

The fact that we predict the computational possibility of unattested forms is not only possible, but highly likely, given the fact that the language faculty is embedded in a complex system of other cognitive and physiological modules with which it interfaces. Consider the following example. Suppose that the rule $\mathcal{R}$ of a formal system combines the primitive categories of the system $\{a, b, c, d, e\}$ into ordered pairs such as $< a, b >$, $< e, c >$, $< b, d >$, etc. Suppose that after collecting a sample of data we notice that all ordered pairs have occurred except for $< a, d >$. If we then supplement our characterization of the formal system by adding a constraint $^*< a, d >$, what have we gained? We have merely built the descriptive generalization into the grammar. Two preferable alternatives come to mind.

The alternative suggested by Pylyshyn's example is to look outside of the formal system itself. In phonology, for example, the shape of phoneme inventories reflects the nature of sound change and physiological constraints on articulation, not just the cognitive capacity of humans. Not only is it misleading and uninsightful to posit constraints on the formal system that do no more than recapitulate observation, but it also discourages us from looking for a real explanation in a domain other than the characterization of the formal system. (see Hale and Reiss 2000*a, b* for discussion.) This approach is adopted by Reiss (2003a) to account for unattested patterns of quantification in phonological rules.[9]

A second alternative to explore is to examine whether $\mathcal{R}$ has been correctly formulated. Many constraint-based linguistic analyses are built by positing a spurious generalization, then adding constraints to the model to account for the cases which do not match the generalization. It seems more elegant to posit our generalizations more carefully. This approach is taken below in our discussion of so-called OCP effects.

Has the preceding dismissal of concerns of overgeneration made the proposals here vacuous? For example, does the position reduce to the following: "posit a rule that generates all the attested data, and assume that unattested data is the result of accidental gaps in the corpus"? Fortunately, the answer is that this is not the position I am advocating, and this is because of a simple claim that is in direct conflict with general practice, at least in the phonology literature. The claim is that rules are formulated in the *least* general form that is compatible with the data.[10] Generality of application results from lack of specification in structural descriptions; lack of generality, that is, restrictiveness of application results from richly specified structural descriptions. In the

[9] A paper by Bakovic (2005) which criticizes several aspects of this discussion came to my attention as this chapter was going to press. Unfortunately, the careful response that Bakovic's paper deserves cannot be undertaken here.

[10] For example, a palatalization rule that applies before the vowels [i, e] in a language with only the vowels [i, e, a, u, o] should be formulated with the conditioning environment as "before [−back, −round, +tense, −low] vowels", and *not* as "before [−back] vowels".

view of acquisition developed in Reiss (1995, 1999) and Hale and Reiss (2003), it is claimed that representations that are more highly specified than necessary for the purposes of generating target output, are a logical necessity in early grammars. Rules are only made more general, that is with less specified structural descriptions, upon exposure to positive evidence. Therefore, a rule of a particular grammar will generate all and only the data whose representations are subsumed by that encountered during the acquisition process.

8.5 A right-minded approach to syntax

The conclusion to be drawn from the discussion above is that it is in fact best to state our theory of UG in terms of a positive list of what can occur—what Wolfram would perhaps call a list of possible components for programs. This approach actually does delimit the set of possible languages as well as a theory that states constraints on possible linguistic structures, because the normal interpretation of a formal system defined by a set of properties (a vocabulary) is that the system is exhaustively defined by those properties. (See Rennison (2000) for an explicit discussion along these lines.) One can add or subtract one of Euclid's Postulates and explore the consequences of such a move, but any set of postulates is assumed to be exhaustive once stated. Similarly, in physics, new elementary particles are posited only when a phenomenon cannot be accounted for by appeal to those currently identified, or when their existence is predicted on other grounds. Since linguistics posits formal models of (indirectly) observable systems, our current theory is open to revision when forced by new discoveries, but Occam's Razor serves as a check on the current version at any particular time. A model characterized by prohibitions in the form of constraints must implicitly be itself constrained by a vocabulary defining the universe of discourse in which the constraints hold. Therefore, such a model contains a certain amount of unnecessary redundancy.

The derivational approach to syntactic relations developed in Epstein, Groat, Kawashima, and Kitahara (1998) adopts a viewpoint consistent with the "rules only" approach to modeling grammar advocated here. These authors claim (pp. 13–14) that their theory has five innovative properties. The first and the last are most clearly relevant to the discussion in this chapter and can be summarized as follows:

(6) Epstein, Groat, Kawashima, and Kitahara (1998)

- The syntactic computational system consists only of syntactic rules. There are no relations (like *Government*) that are not derivable from the nature of the rules;

- There are no filters or constraints (on non-existent levels of representation such as DS and SS), but only lexical items and operations on these items.

These authors are able, for the most part, to do away with independently stipulated constraints on movement such as GREED and SHORTEST MOVE and instead build their effects into the nature of the rule/process *Merge* itself. I understand the goal of this model to be to formulate a rule/process *Merge* which applies in such a way that its outputs are well-formed, as long as it is possible to generate a well-formed output from the current input. Perhaps a better way to describe the model is to say that outputs are "formed", or "not formed", and that the notion "well-formed" is undefined—and unnecessary.

In the rest of this chapter, I explore a parallel approach to phonological derivation. First, I provide some background on the use of constraints within primarily rule-based phonologies. Then I demonstrate the insight that can be gained by building the effects of constraints into the statements of the rules themselves.

8.6 Constraints in rule-based phonology

Despite the fact that phonologists tend to characterize current debate concerning OT as a question of "rules vs. constraints", this is misleading (see Archangeli 1997). Many rule-based analyses make use of constraints such as the Obligatory Contour Principle (OCP). Constraints in otherwise rule-based phonologies serve two main purposes. Either they define certain structures as disfavored or ill-formed, and thus subject to modification by rule; or they are used to block the application of a rule just in case the rules' output would be disfavored or ill-formed. Work by Paradis (1988*a*) and Calabrese (1988) are typical of the use of constraints as diagnostics for repair of certain structures: if a string satisfies the structural description of a constraint, that is, if it violates the constraint, it must be repaired by a rule. The rule-based account of stress systems presented by Halle and Idsardi (1995) appeals to "Avoidance Constraints" (422ff.) which prevent the application of rules in cases where the rules' output would be a "disfavored" structure. The OCP has been invoked for both of these purposes in a number of papers, most notably McCarthy (1986) and Yip (1988).

Given the problems with markedness theory alluded to above, note that in the absence of a theory of disfavoredness, this approach is circular: the only real evidence for the disfavored status is that the posited rule appears to be blocked; and the posited reason for the blocking is that the resultant structure would be disfavored. Halle and Idsardi point out that certain advantages derive from mixing rules with constraints in the analysis of individual languages. In general, the use of constraints allows us to formulate simpler rules. However, they note that a fully rule-based analysis is in principle always possible—Halle and Vergnaud (1987) is an example they cite:

> In Halle & Vergnaud (1987), the full metrical constituency was constructed, and at the end disfavored configurations [like stress clash] were eliminated by the application of a rule.

I propose that considerations of elegance for a theory of UG take precedence over elegance in the analysis of individual languages, and thus the Halle and Idsardi system, for example, should be adapted in a way that preserves its mathematical explicitness, while doing away with constraints on unattested structures. A possibility which Halle and Idsardi do not consider[11] is to make the structural descriptions of their rules more complex. As these authors point out, some languages do tolerate stress clash and thus their avoidance constraint is specific to those languages which do not tolerate clash. The rewards of allowing for more complex rules are considerable: constraints become unnecessary and the effects of earlier rules need not be undone.

In brief, Halle and Idsardi need the avoidance constraint Avoid(x(to prevent the generation of Line 0 metrical structures such as (x (x x (x x in a language like Garawa that (1) inserts the leftmost left parenthesis on the basis of an Edge-marking rule, and (2) inserts left parentheses iteratively from the right edge after every second syllable. In a word with an even number of syllables, steps (1) and (2) give, e.g., *(watjim(paŋu*. However, in a word with an odd number of syllables the rules outlined above would generate a "disfavored" (x(structure like *(na(řiŋin(muku(njinam(iřa* where the leftmost syllable has a left parenthesis on both its right and its left. The avoidance constraint blocks the insertion of a parenthesis to the left of the second syllable from the left, and the actually generated Line 0 form is *(nařiŋin(muku(njinam(iřa* with a trisyllabic leftmost constituent. Instead of appealing to an avoidance constraint, the so-called Iterative Constituent Construction rule can be specified to insert a left parenthesis only in the environment x x _ x x. By the normal conventions of interpretation, the structural description is not satisfied by the following structure: x (x _ x x. Thus, the stress-clash configuration is not generated.[12] Again, we cannot rule out such complications to rules *a priori*, without

[11] Idsardi (1992), however, does have a useful discussion of rule-, constraint-, and rule-and-constraint-based approaches to stress.

[12] Because it is not relevant to the discussion, I ignore here the further steps in the derivation, those which follow the construction of the Line 0 structure.

considering that the use of the simpler rule requires adding an additional rule to the grammar (in the Halle and Vergnaud formulation) or else enriching grammatical theory by the use of avoidance constraints (in the Halle and Idsardi formulation).[13]

I thus propose that a goal of future phonological research should be to take the idea of rule-based phonology seriously—by avoiding constraints altogether. Such an approach will offer a principled alternative to Optimality Theory and other constraint-based models. In other words, rather than stating simple, but empirically inadequate rules, reinforced by an arsenal of language-particular or universal constraints, we should attempt to understand what kind of rules we actually need if we are to do without any constraints.

Part of the groundwork for this approach was done over ten years ago in a pair of underappreciated papers by David Odden (1986, 1988). Odden demonstrated that the OCP is demonstrably *not* a universal constraint on either underlying representations or on the workings of the phonological component. Odden also points out that work appealing to the OCP is unacceptably vague in defining how, for example, identity of representations is computed. These arguments need not be repeated here, since my goal is to reject the use of all constraints on more general grounds.

8.7 The Obligatory Contour Principle

McCarthy (1986) discusses data from several languages in which a vowel which is expected for independent reasons to be deleted, is instead preserved if its deletion would cause identical consonants to be adjacent: Biblical Hebrew /ka:tab-u:/ → [ka:θvu:] but /sa:bab-u:/ → [sa:vavu:] because deletion would bring together the two underlying [b]s (both of which are spirantized by an unrelated process).[14] The "failure" of the deletion rule to apply is dubbed *antigemination* by McCarthy, since the rule is "blocked" if its application would produce a geminate. McCarthy invokes the Obligatory Contour Principle (OCP) as the constraint which blocks the rule from applying. This

[13] There are, in fact, other plausible rule-based analyses. Morris Halle (pers. comm.) points out that by first building a single binary foot from the *left* edge of the word, then building binary feet iteratively from the right, the third syllable from the left will remain unfooted in words with an odd number of syllables, but not in those with an even number.

Even number of syllables: x x) (x x (x x
Odd number of syllables: x x) x (x x (x x

By projecting the leftmost syllable of each foot, the correct Line 1 configuration is generated for all words.

[14] It has been brought to my attention that vowel length in Hebrew is actually difficult to determine. However, this issue is irrelevant to the point under discussion—any example of "antigemination" will do and additional ones are provided below.

phenomenon involves the failure of deletion rules just in cases where the rule would result in a string of identical adjacent consonants.

Yip (1988) provides a very useful summary, elaboration, and discussion of McCarthy's treatment of the OCP as a blocker of rules. Consider the following argument:

> If a language has a general phonological rule that is blocked just when the output would contain a sequence of identical feature matrices, we can conclude that the OCP is operating to constrain derivations ... The alternative is an *ad hoc* condition on such rules, as in [7]:
>
> (7) $A \rightarrow \emptyset / B \_\_ C$
> Condition: $B \neq C$
>
> Such a condition not only incurs an additional cost (whereas the OCP is taken to be universal) but also lacks explanatory power, particularly if contexts B and C are necessary only to state the *ad hoc* condition.

In other words, Yip argues that a theory with language-specific rules and a universal OCP is a better theory than one with language-specific rules that correctly encode where the rule applies, because adding the necessary conditions to the statement of such rules makes them more complex.

Note that the examples that Yip mentions conform to the first (a) of the following three types of conditions on rule application, but Odden (1988) points out that in fact vowel syncope rules are found with all three of the following types of conditioning:

(8) Some conditions on vowel deletion rules (Odden 1988: 462)

a. Delete a vowel unless flanking Cs are identical.

b. Delete a vowel blindly [whatever the flanking Cs are].

c. Delete a vowel only if flanking Cs are identical.

Condition (a) can be restated as "Delete a vowel if flanking Cs are *not* identical." This is the condition described but rejected by Yip in (7) above: $B \neq C$. But note that Odden's type (c) condition would be written as follows:

(9) Odden's condition (c) in the notation Yip rejects: $B = C$

In other words (a) demands non-identity and (c) demands identity of segments in the structural description of a rule. Thus, there is no reason to propose, as McCarthy and Yip do, that rules that conform to condition (a) illustrate a universal principle of markedness—condition (c) is also a possible rule condition. A rule like (8c) *only* applies when it creates OCP violations—Odden refers to this phenomenon as *antiantigemination*. So a theory of UG must allow for both types. There is thus no good reason to claim that a universal principle, the OCP, *blocks* deletion in the (a) cases, since deletion can also be *required* in cases that lead to apparent OCP violations when a rule with conditions (b) or (c) applies. Stated in McCarthy's terms (although he does not mention such cases), deletion can be blocked (in case (c)) if the rule will *not* generate an OCP violation. This point was clearly made by Odden, though it seems to have been ignored in most of the subsequent literature.[15]

Note that the logic of attributing cases that fit the profile of (a) to a universal principle and ignoring cases that fit (c), is incoherent. Suppose we examine some data concerning a certain phenomenon and find that all cases fall into two categories, *x* or *y*. If we present only cases of *x* and proclaim that we have found that *x* is always true, then our claim is not valid, *no matter how many positive examples of x we adduce*. The existence of (c) cases makes the existence of (a) cases uninteresting on their own. Odden's observations taken together *are* interesting, as we will see below. Simply put, case (c) is a counterexample to the claim that (a) is universal.[16]

8.7.1 *Treating phonological pathology: The OCP as a rule trigger*

The main point of Yip's paper is that the OCP not only *blocks* rule application as in McCarthy's antigemination cases, but also *triggers* it—it may be the case that a rule applies only to an input that violates the OCP. Instead of an argument based on formal simplicity in rule statements, as discussed above, Yip's discussion of the OCP as a rule trigger illustrates particularly well the assumption that the phonology repairs structures that are somehow pathological—ill-formed or marked or disfavored: "The main contribution of the OCP is that it allows us to separate out condition and cure. The OCP is a trigger, a pressure for change" (74).

In Yip's model the "cure" is effected by language-specific rules. In OT models that make use of similar constraints the "cure" emerges from the constraint ranking. Because of the violability of OT constraints, the winning candidate in an OT derivation is typically not fully "cured'—certain marked structures may be present in the output form.[17] One goal of this chapter is to work towards removing the notion of ill-formedness from the generative component of the phonology. There are representations that are generated,

[15] For example, Keer's (1999) recent OT thesis on the OCP, lists Odden's papers in the bibliography, but makes no reference to them in the text, even in sections discussing antigemination.

[16] Providing a principled response to the reader who finds this discussion to constitute an argument for the violable constraints of Optimality Theory is beyond the scope of this chapter, or perhaps even impossible, reducing to a question of faith.

[17] We might refer to this idea as OT's Fallacy of Imperfection. Imperfection, or markedness, seems to be as irrelevant to linguistic theory as the notion of perfection.

or formed, by grammars; there are representations that are not generated—that is, not formed; but there is no reason to believe that anything a grammar actually generates is ill-formed.

Yip provides a range of examples that show how different solutions can be applied to OCP violations. They include deletion, dissimilation, and assimilation rules (where assimilation represents multiple linking of a single node, and not identical adjacent nodes). One example of repair by deletion comes from Seri (Marlett and Stemberger 1983). This language has a rule that deletes a coda glottal stop in a syllable with a glottal stop in the onset:

(10) *Seri Glottal Stops*

a. ʔa-a:ʔ-sanx → ʔ-a:-sanx 'who was carried'

b. ʔi-ʔ-a:ʔ-kašni → ʔi-ʔ-a:-kašni 'my being bitten'

c. koʔpanšx 'run like him!'

The rule only applies to tautosyllabic glottal stops so the second glottal stop in (10b) is not affected. In general, coda glottal stops can surface, as shown by (10c).

Yip's account of this process is the following:

> [We can] assume that the Laryngeal node is absent except for /ʔ/, and the entries for glottalization in [10ab] are thus adjacent and identical and violate the OCP. This violation triggers a rule that operates in the domain of the syllable, and the language chooses [one of the possibilities for repairing OCP violations,] deletion of one matrix (either [+constricted] or [Laryngeal]). The actual rule has four parts, as shown in (11):
>
> (11) *Glottal Degemination*
> Domain: Syllable
> Tier: Laryngeal
> Trigger:
> Change: Delete second
>
> The environment is not stated, so the rule is unable to operate unless triggered "from the outside". The outside trigger is, of course, the OCP, a universal principle and thus free of charge.

In another example, Yip proposes that English uses epenthesis to "cure" OCP violations of adjacent coronal stridents, thus accounting, for example, for the form of the plural morpheme after coronal stridents: *judges, couches, bushes, cases*, etc. In other words, if epenthesis did not apply, the adjacent coronal stridents would constitute an OCP violation. As Odden (1988) points out, the OCP is invoked rather opportunistically—note that it appears to be irrelevant to identity of adjacent [+voiced] specifications in words like *bins, rugs, hills,* *cars.* More seriously, Odden points out that there are rules that insert vowels only when doing so will specifically *not* repair an OCP violation. This is case (d) below. There are also rules that insert vowels regardless of the nature of the flanking consonants—case (e). And of course, there are rules that, like English epenthesis, depend on the total or partial identity of flanking segments—case (f).

(12) More conditions on vowel insertion rules (Odden 1988: 462)

d. Insert a vowel unless flanking Cs are identical.

e. Insert a vowel blindly [whatever the flanking Cs are].

f. Insert a vowel only if flanking Cs are identical.

Parallel to (a), condition (d) can be restated as "Insert a vowel if flanking Cs are *not* identical." Thus there is no reason to see (f) as reflecting the OCP as a trigger when (d) shows that rules may be triggered if and only if they *fail* to fix OCP violations. The existence of rules with conditions (c) and (d) make it unlikely that appealing to the OCP as either a trigger or blocker of rules is a fruitful endeavor.

8.7.2 *The* IDENTITY *and* NON-IDENTITY CONDITIONS

More of Odden's data will be presented below. For now, note that it is equally possible for a rule to generate OCP violations (c) as it is to repair them (f). And it is equally possible for a rule to be "blocked" from generating OCP violations (a) as to be blocked from fixing them (d).[18] Since the goal of phonological theory should be to define the set of computationally possible human languages, Odden's observations provide an excellent opportunity to study the purely formal nature of linguistic rules. In the following discussion, we will concentrate on syncope rules as a matter of expository convenience. Again, for expository convenience, we will refer to a schematic representation C_1VC_2. Odden's conditions (a) and (c) can be restated in the following:

(13) The NON-IDENTITY CONDITION on syncope rules (Version 1)
Delete a vowel if flanking Cs are *not* identical ($C_1 \neq C_2$).

(14) The IDENTITY CONDITION on syncope rules (Version 1)
Delete a vowel if flanking Cs are identical ($C_1 = C_2$).

The apparatus of phonological representation must be at least powerful enough to express the NON-IDENTITY CONDITION and the IDENTITY

[18] Of course, (b) also potentially generates OCP violations, and (e) potentially repairs OCP violations.

CONDITION. This issue has implications for feature geometry as a model of phonological representation. There is an insightful discussion of the need for Identity Conditions in Archangeli and Pulleyblank (1994: 368–73). These authors point out that "linked structures themselves are simply one type of configuration involving identity" (369). Archangeli and Pulleyblank present the "Identity Predicate," a relation holding between two arguments, which "is important in a wide variety of phonological contexts" (369). In addition to the OCP cases, they cite the case of Tiv where [+round] spreads between vowels, if and only if they agree in height. Arguments against a linked structure analysis of identity conditions include cases where identity holds across a morpheme boundary—since the identical features belong to different lexical items, they cannot be stored as linked.

In Reiss (2003*a*), I formalize the identity and non-identity conditions and offer further arguments for the inadequacy of a "linked structure" analysis of these conditions. I also argue that autosegmental feature geometry cannot express such conditions, and that a sufficiently powerful formalism makes feature geometry unnecessary, and thus not part of phonological theory.

The crux of the argument against autosegmental representation is that non-identity conditions require that two segments be distinct. This cannot be expressed using just feature geometric association lines. For example, imagine a requirement that C_1 and C_2 be different with respect to some arbitrary feature, that is any feature, or any feature out of a predefined subset of all the features. In other words, the two segments must *not be identical*, but it doesn't matter how they differ. In order to express such a NON-IDENTITY CONDITION we can make use of something like the existential quantifier: there exists at least one feature for which C_1 and C_2 have different values.

8.8 Constraints alone vs. Rules and Constraints vs. Rules alone

A reader may have been convinced by this brief sketch to accept the necessity for the additional power granted to the representational component argued for here—the necessity of quantification—without accepting rejection of constraints. The formulation of constraints that can evaluate identity and non-identity would also require the use of quantification. Therefore, constraints on their own, or constraints in conjunction with rules do not vitiate the need for quantificational statements in grammars.

Consider, however, what we gain by adopting a minimalist approach to characterizing the phonological component in terms of rules: we have a rule component which allows the use of quantificational statements; we have no notion of well-formedness or ill-formedness—the phonology maps inputs to outputs. In the following table I compare three approaches to building a phonology, under the assumption that they are all empirically non-distinct, that is, that they can generate the same sets of output. The Just Rules (JR) approach outlined in this chapter is compared to "standard" OT and a generic Rules and Constraints (RC) model.

(15) Comparison of various approaches to phonology

	OT	RC	JR
a. List of Primitive Entities	yes	yes	yes
b. List of Possible Operations/Functions	yes	yes	yes
c. List of Constraints	yes	yes	no
d. Notion of Ill-formedness	yes	yes	no
e. Notion of Repair	no	yes	no
f. Quantifiers in SDs	yes	yes	yes
g. Representational Matching Procedure	yes	yes	yes

A complete formal theory of phonology must specify what it can generate, so it is necessary to define the universe of discourse by listing the entities (a) and operations (b) that the computations have access to. In OT there are no rules, but as discussed above, a fully explicit version of OT will have to provide a finite characterization of what Gen actually does—a list of possible operations on representations is in fact a necessary part of the model. In addition, OT contains other functions, such as Eval, so all three theories contain functions. The three models cannot be distinguished on these grounds.

Obviously, there are constraints (c) in OT and RC models, and there are none in JR. As Yip explains, the use of constraints presupposes a notion of ill-formedness (d), which I have argued is circular at best, and incoherent at worst, as an explanation of phonological alternation. The constraints are posited on the basis of this intuited sense of well-formedness vs. ill-formedness or markedness. This notion does not exist in the JR model, in which a set of rules maps phonological inputs to outputs.

OT does not prescribe a specific repair (e) for individual markedness violations, but conceives of the grammar as finding an optimal solution across all outputs, which emerges from the ranking. In RC, rules are applied to repair ill-formed structures or to block rule application, thus also appealing to markedness theory. Repair is not part of JR theory. In all three theories, quantifiers (f) are necessary to evaluate the SDs of rules or constraints which refer to identity and non-identity. Similarly, all three theories need some kind of Representational Mapping Procedure to determine which representations satisfy the structural description of its rules or constraints.

Recall that we are assuming that we can compare extensionally equivalent grammars. While straightforward theory comparison is difficult, the "rules only" approach appears to be the most elegant. The list of possible operations is stated in positive terms and thus characterizes the universe of discourse with no additional apparatus. There is no notion of markedness, and thus no reason to conceive of rules as repairing representations. The theory requires rules with a sufficiently rich representational apparatus to define their condition of application. However, as exemplified by the discussion of quantification, this apparatus may be needed by any empirically adequate theory.

8.8.1 *Violability and universality in Optimality Theory*

Optimality Theory is a model of grammar which posits universal, violable constraints that are ranked on a language-particular basis. The universality and the violability of OT constraints are not independent. Obviously, different constraints appear to hold in different languages, so if constraints are universal, they must be violable.

One might maintain an OT-type computational system of ranked constraints while denying the universality of constraints. However, if constraints are not universal, then they must be learned for each language. If they are learned, then they could be learned with appropriate structural descriptions that make them surface-true (ignoring the possibly insurmountable problem of opacity for two-level theories like OT). If they are surface-true, then they need not be violable. In other words, if we weaken the claim to universality of OT constraints the rest of the theoretical edifice of OT crumbles as well.

8.8.2 *Structural descriptions are "constraints" on application*

Let's look back to the type of rule discussed by McCarthy to motivate the restriction of rule application by the Ocp. Notice that blocking of a rule R can be achieved in one of two ways—either by applying R and undoing its effects if they are "undesirable," or by "looking ahead" to see what the output would be before applying R, and not applying R if the projected output is undesirable. There is, however, a simpler way of avoiding rule outputs that result in ungrammatical surface forms: reformulate the rule as R', so as to apply only when it should. We have said this much already, however, it is important to realize that the structural description of a rule, the representation that determines whether the rule applies via the representational matching procedure discussed in Section 8.3.1, is nothing other than a constraint on application. McCarthy's rule of vowel syncope in Hebrew applies to vowels between consonants, not to any segment that is between any other two segments. The rule applies only under certain metrical conditions, not under others. The condition that the flanking consonants be non-identical, is thus of the same type as the other constraints on application, the other components of the rule's structural description. In other words, there is no motivation in a rule-based grammar that uses an RMP to also have constraints that are not just part of the structural description of rules.

Analogies may again be useful. There is no reason to assume that a law of Newtonian physics, $f = ma$, that refers to entities like *force*, *mass*, and *acceleration* is actually better seen as a relation between variables $x = yz$, which is constrained by a constraint system that rules out any possible instantiation of $x = yz$ other than $f = ma$. Similarly, a rule or law includes a specification of when it is applicable. Writing highly general rules that lack appropriate structural descriptions to restrict sufficiently when the rules actually apply, and then positing constraints that limit the applicability of a rule seems unproductive.

8.8.3 *What is a possible rule?*

Recall that Yip claims that the fact that OCP "effects" are quite common in the languages of the world should motivate us to remove identity and non-identity conditions from structural descriptions. I suggest that this is exactly the wrong conclusion. These types of conditions are among the most crucial things we need to understand if we want to understand how to characterize the class of possible phonological rules. Ironically, such important empirical work by Yip and McCarthy led to the rejection of rule-based phonology in favor of OT, when it should, instead, have led to a deepening of our understanding of the nature of phonological rules. By appealing to constraints we complicate the theory of grammar unnecessarily, since the RMP used in the structural description of rules already provides the computational power that additional constraints were meant to supply. In addition to this complication we also obscure the question "What is a possible rule?"

8.8.4 *What is Universal Grammar?*

A common characterization of the content of a theory of universal grammar presents the goal of UG theorizing to be a search for properties found in all languages. OT in some sense has solved the problem of UG, thus formulated. All constraints are assumed to be present in all languages; however, because some constraints are outranked by conflicting ones, the effects of the former may not be visible in a particular grammar. For example, all grammars have a constraint FAITHSUC demanding input-output faithfulness for the feature

Suction associated wth clicks. However, in English, it is assumed, the markedness constraint NoSuc outranks the faithfulness constraint, so that clicks would not surface, even if they appeared in an English input representation. Thus we see no evidence for FaithSuc by examining English.

Unfortunately, this approach to universalism seriously misconstrues the nature of theorizing about UG since Chomsky's earliest work. The issue is even discussed as early as Lyons (1970):

(16) Lyons (1970) on Chomskyan UG

- "languages make use of the same formal operations" (p. 115).
- "Chomsky believes that there are certain ... units that are *universal*, not in the sense that they are necessarily present in all languages, but in the somewhat different and perhaps less usual, sense of the term 'universal,' that they can be defined independently of their occurrence in any particular language and can be identified, when they do occur in particular languages, on the basis of their definition within the general theory" (p. 111).
- "Chomsky accepts that any one of his allegedly universal features might be absent, not only 'from the very next language that becomes accessible,' but also from very many quite familiar languages" (p. 114–15).

Another angle on the Chomskyan view recognizes UG, not as a hypothesis, but as a topic of study, the study of the initial state of the language faculty: "In any computational theory, "learning" can consist only of creating novel combinations of primitives already innately available" (Jackendoff 1990: 40; see also Fodor 1976 and Pylyshyn 1973). Therefore, the OT approach to universalism, which attempts to reduce all language variation to constraint ranking follows from an overly simplistic conception of what UG is. By ascribing all constraints to all languages, OT has solved a problem that derives from a misunderstanding of the nature of the enterprise of UG: "How can we define the 'units' that are present in all languages?" This is a different problem from determining the nature of the human language faculty.

8.9 Constraining learning

The preceding sections were devoted to a critique of the use of constraints in linguistic theorizing. One argument depended on the idea that constraints are epiphenomena that must be interpreted in the context of a more or less explicit characterization of the universe of discourse in which they apply.

A second argument depended on a rejection of the notion of markedness or ill-formedness, on which many constraints are based. We now turn to discussion of how the language learner's search space can be constrained or limited, without recourse to positing constraints as a component of Universal Grammar.

Consider a simple case of allophonic patterning such as the distribution of light and dark laterals in Georgian (Robins and Waterson 1952). The light [l] occurs only before front vowels and the dark [ɫ] occurs elsewhere. The language has five vowels [i,e,u,o,a] so we have several options concerning how to formulate the relevant rule. Let's consider two of them. We could either formulate a rule that said "/ɫ/ > /l/ before non-low front vowels" or a rule that said "/ɫ/ > /l/ before front vowels":

(17) Georgian lateral fronting

- Vowels: [i,e,u,o,a]
- /ɫ/ > [l] before i and e

a. $\begin{bmatrix} \text{+lateral} \\ \text{+son} \\ \vdots \end{bmatrix} > [-\text{back}] \text{ before } \begin{bmatrix} -\text{back} \\ +\text{ATR} \\ -\text{low} \\ -\text{round} \end{bmatrix}$

OR

b. $\begin{bmatrix} \text{+lateral} \\ \text{+son} \\ \vdots \end{bmatrix} > [-\text{back}] \text{ before } \begin{bmatrix} -\text{back} \end{bmatrix}$

No language-internal evidence would bear on the matter of selecting the correct formulation of the rule since the only front vowels in the language are non-low. In other words the rules are extensionally equivalent. Despite the fact that we tend to teach beginning students that the second rule is better, since it is more concise, I will argue that the first is the better solution.[19] In the course of this discussion, I hope to convince you that you should care which answer is closer to the correct one, and that justifying this claim is not as hopeless an enterprise as it has seemed to be in the past.

8.9.1 *Approaches to phonology*

Suppose that a child, Junior, is acquiring an "English-type" grammar, the output of which includes forms like [k^hæt]. It seems clear that cognitive scientists,

[19] In Reiss 2003*b* I provided arguments based on cross-linguistic patterns against choosing the most concise formulation of a rule. Here I offer a different kind of evidence leading to a similar conclusion.

phonologists in particular, should set as an ultimate goal finding a solution to the first of the following questions (which is the harder and more interesting one), and they should not be satisfied with merely answering the second.

(18) Two kinds of question

- 'What knowledge state underlies Junior's output such that he says [kʰæt]?'
- 'What is the set of possible knowledge states that could lead to Junior saying [kʰæt]?'

The answer to the first question correctly entails a concern with I-language, language conceived of as knowledge, a matter of "individual psychology" (Chomsky 1986). In other words, phonology is computation over symbolic representations by the phonological component of the mind/brain. Let's refer to this approach as the I-phonology approach. The second is merely concerned with defining extensionally equivalent E-languages, that is language conceived of as sets (or potential sets) of utterances, tokens of behavior. This "E-phonology" approach may involve some interesting theorizing on the formal properties of grammars, both humanly attainable ones and others; however, it cannot be adopted as the right approach for phonology as cognitive science.

I will argue that much of the phonological literature, both before and since the advent of OT, has given up on answering questions of the first type. In fact phonologists have turned away from this goal in at least two ways. "E-phonologists" are concerned with the formal issues entailed by the second type of question. It is important to note, however, that like "I-phonologists" they are concerned with mappings between input and output *representations*.

Others have turned further from the goal of I-phonology in their sometimes tacit rejection of the supposition that phonology is only about knowledge and representations. Instead this work is concerned with more superficial,[20] data-fitting theories of speech output as *behavior*. We can thus refer to this as the "B-phonology" school.

We can characterize the concerns of the three-way distinction we now have with these questions:

(19) Three approaches to phonology

- I-phonology: 'Which humanly attainable knowledge state underlies Junior's computation over phonological representations?'

[20] This word is meant in the sense of "observable," not in a necessarily pejorative sense, although I do believe the approach is misguided.

- E-phonology: 'What is the set of formal systems that would output the same representations as Junior's phonology outputs?'[21]
- B-phonology: 'What can we say about the *sounds* Junior makes?'

The "evaluation procedures" discussed in *SPE* and subsequent work were meant to answer questions of the first type, but Anderson's (1985: 327) remarks on the topic are telling: "Early concern for evaluation procedures…turned out to be something of a dead end." and "[T]he appeal of feature counting went away…not with a bang, but with a whimper." I will discuss some simple examples which suggest that prospects for answering the first type of question are not as bleak as they have seemed in the past, and that I-phonology is thus a viable enterprise. I thus attempt to revive these issues by redefining the relationship between the study of phonological theory *per se* and phonological acquisition and learnability.

In addition to making positive proposals, I will point out where other models of phonology have strayed from the pursuit of I-phonology. With respect to OT in particular, I will argue here that the notion of *Richness of the Base* has no place in a theory of I-phonology, and that endowing the learner with an innate set of constraints referring to phonetic substance does nothing to solve the paradox of language acquisition.

8.9.2 *Two reasons to look at acquisition*

Given Kiparsky's (1973: 17) observation that "Children learning their native language do not have the interests of linguists at heart" it is necessary that we view phonology from the learner's perspective. Our reward for such attention to the acquisition process will be twofold. First of all, paying attention to acquisition can tell us what we need *not* worry about. For example, the OT literature is rife with claims of OT's superiority at accounting for conspiracies: "One of the principal reasons that rule-based theory has come under attack is that it offers no satisfactory explanation for conspiracies" (Kager 1997: 463). Kiparsky (1973) has shown, however, that generative phonology does not need the notion of conspiracy. Here is my interpretation of Kiparsky's argument.

[21] There is yet another possible subdistinction: some E-phonologists might concern themselves with only humanly attainable formal systems. I will argue in this chapter that it is useful to assume that given the assumed invariance of the language faculty, only one grammar is attainable on exposure to a given set of input data.

(20) The epiphenomenality of conspiracies (based on Kiparsky 1973: 75ff.)

i. A conspiracy is a set of rules that are "functionally related," that is they lead to the same kinds of output configurations such as "all syllables are open."
ii. If a language has such a set of rules, then the rules of the language will tend to be surface-true (transparent).
iii. Non-transparent (opaque) rules are not surface-true.
iv. Rules that are not surface-true are hard for a learner to learn.
v. Things that are hard to learn are more likely *not* to be learned than things which are easy to learn.
vi. Failure to learn aspects of the ambient language constitutes a diachronic change.
vii. Therefore, (E-)languages are more likely to lose opacity than gain opacity.
viii. Therefore, grammars are likely to look like they have conspiracies.

In other words, the existence of conspiracies is an epiphenomenon due to the fact that languages tend to have transparent rules. This in turn is an epiphenomenon derived from the undeniable fact that individual languages must be learned.

Kiparsky's explanation of conspiracies depends on the fact that acquisition can be *unsuccessful*, resulting in so-called language change (Hale 2007). In other words, tendencies such as "conspiracies" are to be explained by reference to diachronic linguistics where the goal is to define possible changes and to explain why certain changes are more or less likely to occur. We now turn to the question of what *successful* acquisition can potentially tell us.

The second benefit of paying attention to acquisition is that it allows us to take seriously the idea expressed in Chomsky (1986: 3) and elsewhere that Universal Grammar (UG) is the Language Acquisition Device (LAD). In other words, the LAD constrains the set of possible languages by determining how the learner assigns analyses to data provided in the environment, the Primary Linguistic Data (PLD). There are several advantages to such an approach. First, we need no "principles" of UG which are not derivable from, or reducible to, the nature of the LAD. Since we obviously need a learning algorithm (the LAD), a theory with just an LAD is *ceteris paribus* better than a theory with an LAD *and* stipulated principles of UG. This approach also obviates the need for an evaluation metric. Learners never compare extensionally equivalent grammars for simplicity or economy, they just construct the one grammar that is determined by the LAD. This means that there is no reason to introduce the terms "simplicity" and "economy" into the theory since they are contentless labels for aspects of the LAD that are not derivable, that is, they are arbitrary. Note that even if the attempt to collapse UG and the LAD is ultimately misguided, this is not a bad kind of mistake to make. Attempting to collapse the two can lead to the discovery that *some* aspects of our current theory of UG are derivable from the nature of the LAD. Using such findings we can formulate a more streamlined version of UG (*qua* set of stipulated properties of the language faculty not derivable from the LAD) even if we cannot reduce its contents completely.

8.9.3 *Too formal*

In this section I argue, in apparent contradiction to the preceding one, that in some ways, phonologists have been *too* formal in their methods. The contradiction is merely apparent, however, and the problem is mostly one of focus. Since, as Chomsky (1986) puts it, generative lingustics is concerned with matters of "individual psychology" the regularities in the output of linguistic systems need to be seen as the result of innate and learned factors. Focusing on purely formal statements concerning potential *in situ* grammars which are extensionally equivalent misses something critical in that it does not force us to discover *the correct grammar* that constitutes knowledge of some language. Some examples will prove helpful.

Kenstowicz and Kisseberth (1979) provide a useful formulation of Kiparsky's Alternation Condition (AC):

(21) The Alternation Condition (AC) as formulated in Kenstowicz and Kisseberth 1979: 215
Each language has an inventory of segments appearing in underlying representations. Call these segments phonemes. The UR of a morpheme may not contain a phoneme /x/ that is always realized phonetically as identical to the realization of some other phoneme /y/.

We need not worry about which, if any, version of the AC is best, or even if the condition is valid in any form—my point here is one of perspective. If we want to equate UG with the LAD, then, instead of proposing the AC as a principle of UG, we should ask "How does the child set up underlying representations? What is the learning algorithm that is used to capture the apparently real patterns manifested by alternations and distribution of sounds?" Kiparsky (1973) pretty much says this in referring to one version of the AC: "a situation which I termed *absolute neutralization* is either impossible or hard to learn, and should therefore in an explanatory theory of phonology be excluded or specified as not highly valued" (65). The explanatory theory Kiparsky refers to is phonological UG. Once we equate UG and the LAD, Kiparsky's stipulated AC becomes unnecessary. If my suggestion is valid, then it is perhaps unfortunate that later work fails to adopt this position, and the AC is treated as a formal *principle* that constrains grammars (including the lexicon), rather than expressing a generalization about how they are constructed.

It is ironic to note that while a fair amount was written on the AC in the pre-OT era, studies of phonological acquisition posited rules of supposed child phonological systems that violated the AC. For example, kids who do not distinguish [ʃ] from [s] because of a purported rule /ʃ/ > [s] that neutralizes the two, are in blatant violation of the AC. If the AC is conceived as a principle of UG it would be unfortunate if it was violated by children's grammars. A coherent theory that takes acquisition into account will provide a learning algorithm that tells us how underlying representations are generated from the *PLD* (in part by denying the existence of "child phonology rules"—see Hale and Reiss 1998). Therefore such a theory does not need the AC.

Can we relate any of this to OT? One oft-touted property of OT is the notion of richness of the base. Given an appropriate constraint ranking a speaker of English could have any one of a number of forms stored for the lexical item that surfaces as [k^{h}æt]. For example, they could have /k^{h}æt/, /kæt/, or /k!æt/. If, say, constraints against clicks and constraints demanding that voiceless stops be aspirated word-initially are ranked high, then all these inputs would surface as [k^{h}æt]. In other words, the surface inventory is not so much a function of the inputs, but more a result of the ranking. This idea, which is supposed to be as applicable to syntax as it is to phonology, is expressed in discussions of richness of the base in Tesar and Smolensky (1998) and Grimshaw (1997).

The set of possible inputs to the grammars of all languages is the same. The grammatical inventories of languages are defined as the forms appearing in the structural descriptions that emerge from the grammar when it is fed the universal set of all possible inputs. Thus, systematic differences in inventories arise from different constraint rankings, not different inputs. The lexicon of a language is a sample from the inventory of possible inputs; all properties of the lexicon arise indirectly from the grammar, which delimits the inventory from which the lexicon is drawn. There are no morpheme structure constraints on phonological inputs, no lexical parameter that determines whether a language has *pro*.

(Tesar & Smolensky 1998: 252).

We can also see that it is inevitable that light *do* exists in [English], given the constraint rankings.

(Grimshaw 1997: 387).

We must however ask the following: If the inventory is due to the constraint ranking, then what determines the ranking? The answer is obviously that richness of the base expresses exactly the wrong generalization. The inventory present in the ambient language determines the ranking.

Now it is not a problem that OT with richness of the base would allow apparent violations of the AC (by merging all underlying clicks with plain velars for example) since the AC is not part of the theory. However, who, if not phonologists, will be responsible for deciding whether the child has underlying /k^{h}æt/, /kæt/, or /k!æt/? Since we are interested in I-language, we can (and must) ask which is the correct grammar, not just what is the class of extensionally equivalent descriptively adequate grammars. Recall that the two approaches under consideration correspond to the questions we began with in (18). If we believe that our job ends when we can answer the second question, and that the first is not important or perhaps not even coherent, then we will have sided with the anti-mentalism of Quine on the I-/E-language debate (see Chomsky 1986).

We see then that richness of the base, is actually a SYMPTOM of not having an explicit learning algorithm. It represents an abdication of the responsibility of figuring out what the speaker has stored. Of course, one can attempt to provide OT with an explicit learning algorithm, but then richness of the base becomes irrelevant to a characterization of linguistic knowledge. This characterization of the anti-mentalism implicit in many OT analyses is explicit in McCarthy (1999*a*: 6): "with faithfulness bottom-ranked, the choice of input [among three alternatives] doesn't matter, since all map to [the same surface form]. So there is no need to restrict the inputs." McCarthy is confusing the issue of the linguist designing a grammar, *qua* computational system, with the problem of discovering which *mental* grammar the learner acquires.

There is no question of "restricting" the inputs, but rather a question of figuring out which inputs the learner constructs given the observed data. It is something of a perversion of terms to label our hypothesis about what the LAD does a "restriction," when in fact we mean "selection of a uniquely defined choice."

8.9.4 *Innateness and learnability*

In general, I follow Pinker's (1984/96) formulation of orthodox generative views on acquisition, learnability, and innateness.

8.9.4.1 *The Innateness Hypothesis is a misnomer* One of these orthodox ideas is central to our concerns and therefore must be made explicit. This is the view that the Innateness Hypothesis is something of a misnomer. In fact an innate UG is a logical necessity: "In any computational theory, 'learning' can consist only of creating novel combinations of primitives already innately available" (Jackendoff 1990: 40; cf. Pylyshyn 1973: 33; Fodor 1976; Hale and Reiss 2003 for detailed arguments). Basically, the idea is that if learners do not have the representational apparatus needed to represent input in a given domain such as language, then they can never develop that apparatus since they won't be able to recognize what they are supposed to be recognizing! (except through maturation, which is a kind of innateness). What this means for our purposes is that the child *must* have initial access to the universal phonological feature set. This view is inconsistent with much work on phonological acquisition, but it is the only view consistent with the logical necessity of innate representational primitives. It is obviously also tacitly accepted in versions of OT that assume an innate universal constraint set. This view is also consistent with well-known results concerning infants' ability to distinguish all possible phonetic contrasts. Empirical and logical considerations thus force us to endow the learner with the full representational apparatus provided by UG.

8.9.4.2 *The Subset Principle* If one adopts the standard assumption that children do not make use of negative evidence in the course of language acquisition, one is thereby married to some version of the Subset Principle. In other words, the lack of negative evidence necessitates the early formulation of decreasingly restrictive hypotheses concerning the target grammar. So, the essence of the Subset Principle is that the initial hypothesis S_0 concerning the target grammar is maximally constrained. Hypotheses are more constrained when they are more specific; and they are more specified when they are formulated with relatively richer representations. In other words, the logic of the Subset Principle converges with the logic of the necessary innateness of primitives and the experimental evidence. We have three independent arguments for initial full access. We see then, that initial representations must be very rich; and that acquisition is a process of "pruning," rather than "growing" structure. This view is relevant to both the representation of lexical items and the representations of components of the computational system—the rules or constraints. I now consider these two subcases in turn.

8.9.5 *Two modest examples*

We can now examine the implications of such "initial full access" for learnabilty and acquisition, and ultimately for the nature of mature grammars.

I hope to show that these proposals lead to some simple new insights. As we will see, the results are not strictly tied to the constraint/rule debate and relate instead to issues of representation.

8.9.5.1 *Underspecification* The issue of specificity (i.e., the choice between rules (17ab)) is closely tied to the issue of underspecification in lexical items: an underspecified phonological representation potentially subsumes more tokens than a fully specified one. It is worth pointing out immediately that the notion of underspecification in adult grammars loses some of its appeal as soon as one recognizes that early grammars could not possibly be underspecified. In other words, in the theory we propose, achieving underspecification requires a longer, not a shorter, learning path for the child as "pruner" than it would for the child as "grower." We recognize several mechanisms for achieving some form of underspecification, but they are all very data-oriented—that is, they are forced by exposure to positive evidence. These mechanisms include the following:

(22) Possible sources of underspecification

- Alternations (see Inkelas 1996)
- Patterns "supported" by alternations: the alternation in *a*[k^h]*use* / *a*[k]*usation* allows the /k/ of non-alternating *cat* to be stored without aspiration (Hale and Reiss 1999*a*)
- Phonetic underspecification as evidenced by gradient transitions (Keating 1988)
- Phonetic underspecification as evidenced by "big target spaces" (Hale and Reiss 2003)
- Transparent segments

This list leaves little motivation for the child (and no empirical evidence for the linguist) to posit radical underspecification or redundancy rules. It is worth pointing out that this argument converges with recent work by Inkelas (1996) and Yip (1994) both of whom reject more "philosophical" approaches to underspecification.

8.9.5.2 *The Generality Problem* In this section and the following one, we turn to consider the degree of specification found in the components of the computational system. For the sake of expository clarity, I illustrate with simple *SPE*-style rules and feature matrices.

A simple formulation of one component of the *SPE* evaluation metric is provided by Kenstowicz and Kisseberth (1979):

(23) The Conciseness Condition (K&K: 336)
If there is more than one possible grammar that can be constructed for a given body of data, choose the grammar that is most concise in terms of the number of feature specifications.

Obviously, a more concise rule or set of rules will be *potentially* more general, just as a less concise (more richly specified) one will be more restrictive. Of course, the Conciseness Condition is intended to compare grammars that are extensionally equivalent, so one might suspect that no empirical evidence could possible bear on the issue of what the correct grammar is for a body of data. However, by forming a hypothesis concerning how the grammar is constructed by a learner, we end up with a hypothesis about its mature form. In other words, talk of feature counting need not be the "dead end" that phonologists once felt it to be if we try to take the learner's perspective.

Two relevant subcases for considering the generality of rules can be referred to as the GENERALITY problem and the VACUOUS APPLICATION problem. Anyone who has taught introductory phonology will be familiar with relevant examples. I will illustrate the generality problem using a comparison between the Georgian lateral fronting mentioned above and English voicing assimilation of /z/. The vacuous application problem will be described briefly.

8.9.5.3 *Georgian lateral fronting* As we saw above, Georgian has a five vowels system containing [i,e,u,o,a]. The language has two surface laterals which are in complementary distribution. Plain or clear [l] occurs before the front vowels [i,e]. The velarized back [ɫ] occurs elsewhere. Therefore, it is relatively straightforward to set up a rule of the form in (24).

(24) Georgian

- Vowels: [i,e,u,o,a]
- /ɫ/ > [l] before i and e

But if we try to formalize this, how general/concise do we make the rule? Should it be stated to apply before [-back] or before [-back, -low] vowels? How could we possibly decide?—No empirical language-internal evidence can tell us since the language has no [-back] vowels that aren't [-low], so we have to rely on what a principled learning algorithm will tell us. Another relevant question is "Why do we care?" The answer is that we get paid to care—phonologists are supposed to explain the nature and content of phonological knowledge, a matter of "individual psychology."

Assuming that I have convinced the reader to care, let's proceed to a demonstration of what an explicit learning algorithm will lead us to. First of all, where does such a rule "come from"? The answer is that it is generated on the basis of some kind of positive evidence, that is on the basis of tokens of the rule's application. Let's gloss over some difficult details and imagine that the learner somehow comes up with the generalization that "/ɫ/ > [l] before i" and also with the generalization that "/ɫ/ > [l] before e." The final rule which is acquired is just the result of generalizing across these two "subrules." This process is achieved, of course, by finding the representation which subsumes the two cases—for our purposes, the intersections of the triggering environment will suffice. An early (i.e., rich, highly specified, restrictive) representation of the two subrules is given in (25):

(25) "Subrules" of lateral fronting

a. /ɫ/ > [l] before i

$$\begin{bmatrix} +\text{lateral} \\ +\text{son} \\ \vdots \end{bmatrix} > [\text{-back}] \text{ before } \begin{bmatrix} +\text{hi} \\ +\textsc{atr} \\ -\text{back} \\ -\text{low} \\ -\text{round} \end{bmatrix}$$

AND

b. /ɫ/ > [l] before e

$$\begin{bmatrix} +\text{lateral} \\ +\text{son} \\ \vdots \end{bmatrix} > [\text{-back}] \text{ before } \begin{bmatrix} -\text{hi} \\ +\textsc{atr} \\ -\text{back} \\ -\text{low} \\ -\text{round} \end{bmatrix}$$

The only generalization (loss of specificity) driven by the data is the pruning of the features where the two subrules disagree. The result is shown in (26).

(26) /ɫ/ > [l] before i AND e

$$\begin{bmatrix} +\text{lateral} \\ +\text{son} \\ \vdots \end{bmatrix} > [\text{-back}] \text{ before } \begin{bmatrix} +\textsc{atr} \\ -\text{back} \\ -\text{low} \\ -\text{round} \end{bmatrix}$$

The representation of the environment thus denotes a natural class that includes both [i] and [e], but not [æ]. Therefore, from an acquisition viewpoint, there is no reason to believe that the child does generalize beyond the data (by choosing a *less specified* statement of the rule). This prediction is testable (with certain caveats), e.g., by testing Georgian speakers' production of lateral-æ sequences.[22]

In order to fully develop the ideas here it would be necessary to present a theory of how morphological parsing is achieved over forms that the learner initially stores in unanalyzed form. However, for the sake of explicitness, we can provide the preliminary statement in (27).

(27) How general are rules? (Formulation 1)
The correct statement of a rule arrived at by the *LAD* is the *most highly specified* representation that subsumes all positive instances of the rule.

There is obviously one thing missing from this formulation, namely a guarantee that it does not overgenerate with respect to attested data. In brief, (27) must be reformulated with a qualification, as in (28).

(28) How general are rules? (Formulation 2)
The correct statement of a rule arrived at by the *LAD* is the *most highly specified* representation that subsumes all positive instances of the rule, and subsumes no negative instances of the rule.

The positive and negative instances of the rule are the stored forms which the learner ultimately parses morphologically in the process of figuring out a phonology and a lexicon. Note that (28) is not a description of what the learning algorithm does, but rather a characterization of the rules it generates. In other words, the representation of the Georgian fronting rule that contains specification that the trigger is [−low] is more highly specified than the representation which excludes the specification of [−low]. The more specific, that is, more restrictive, rule is the one provided by the LAD. Of course, this contradicts the common practice of finding the most economical rule.

Are we to conclude from this that the rules of a grammar are never stated in a form which entails greater generality than that provided by a *list* of positive tokens? The answer, due to the nature of our algorithm, is clearly "no." The result will depend on what representations are subsumed by the acquired representation of the rule. We turn now to a case where the rule is predicted to be more general than what might be predicted *a priori* from a list of positive tokens.

8.9.5.4 *English "overgeneralization"* A standard argument for the existence of phonological rules formulated in terms of features is based on the intuition

[22] If they front laterals before, say, [ɪ], this is not necessarily a problem, since [i] may stand for a vowel which includes the [ɪ] space (see Hale and Reiss 2003).

that English speakers will extend the rule that devoices /z/ after voiceless obstruents to apply even after voiceless obstruents that don't occur in English, such as [x] or [ϕ].[23] In other words, since speakers cannot have memorized that the [-s] form of the plural marker, underlying /-z/, occurs after these sounds, it must be the case that speakers generate the correct, voiceless form on the basis of a rule stated in terms of distinctive features. Let's assume that this intuition is in fact valid and that English speakers do, in fact, pluralize *Bach* as [baxs]. This result is trivially predicted by the learning algorithm which creates rules via subsumption. In (29–30) I have broken down the problem in a manner that is meant to aid exposition, and not to reflect, for example, stages of development. Leaving aside the sibilants, English has the following voiceless obstruents, all of which devoice a following /z/ to [s]: [p, t, k, f, θ]. For simplicity, consider what happens when the contexts of devoicing after various stops are compared. These stops all agree in being [−son], [−cont], [−voice], etc. They disagree in place features such as [ant], [lab], and [cor]. So the representation that subsumes all the stops that trigger devoicing does not contain these place features, but does contain the features for which the stops agree. Note that certain features that are typically assumed to be irrelevant, such as [−lat] are also specified, since there is no mechanism to remove them.

(29) Collapsing place of articulation in stops.

$$\begin{bmatrix} -\text{son} \\ -\text{cont} \\ -\text{voice} \\ +\text{ant} \\ +\text{lab} \\ -\text{cor} \\ \vdots \\ -\text{lat} \\ \vdots \end{bmatrix} \cap \begin{bmatrix} -\text{son} \\ -\text{cont} \\ -\text{voice} \\ +\text{ant} \\ -\text{lab} \\ +\text{cor} \\ \vdots \\ -\text{lat} \\ \vdots \end{bmatrix} \cap \begin{bmatrix} -\text{son} \\ -\text{cont} \\ -\text{voice} \\ -\text{ant} \\ -\text{lab} \\ -\text{cor} \\ \vdots \\ -\text{lat} \\ \vdots \end{bmatrix} = \begin{bmatrix} -\text{son} \\ -\text{cont} \\ -\text{voice} \\ \vdots \\ -\text{lat} \\ \vdots \end{bmatrix}$$

Place features can be similarly factored out across the fricatives, generating a representation which is [−son, +cont, −voice ...] without place features. Finally, the general rule is found by collapsing cases for both stops and fricatives, that is, by eliminating [+/−cont].

[23] Let's keep things simple and not worry about the plurals of words ending with coronal stridents, like *bushes, glasses, beaches.*

(30) The trigger of the devoicing rule

$$\begin{bmatrix} -\text{son} \\ -\text{voice} \\ \vdots \\ -\text{lat} \\ \vdots \end{bmatrix}$$

Since the resultant representation of the triggering environment is [−son, −voice] but not specified for [cont] or place features, this representation describes a natural class that includes [x] and [ɸ]. That is, the most highly specified representation that subsumes the actually occurring cases, also subsumes the plurals of constructed English nouns with final [x] and [ɸ].

To summarize, the LAD constructs a rule *R* whose representation subsumes the description of all positive examples of the rule and no negative ones. Presented with a string (a representation) *S* which is not identical to any previously encountered string, *R* will be appear to be generalized (of course, it is just "applied") if and only if *S* is subsumed by the representation of *R*. This is of course what it means to have a rule. If rules did not work this way, then phonology would not show the kind of productivity that the wug-test manifests.

8.9.5.5 *The Vacuous Application Problem* In Reiss (2003*b*), I argue that the correct rule statement for a process like coda devoicing in Russian or Polish must contain the specification [+voice] in the structural description. In other words, the correct formulation of the rule is more like (31b) than (31a).

(31) Two candidates for the coda devoicing rule

a. [+cons, -son] → [−voiced] in CODA
b. [+voiced, +cons, −son] → [-voiced] in CODA

This conclusion is based on the assumption that the LAD provides a single interpretive procedure for structural descriptions, and that (31a) is the rule a learner formulates on exposure to patterns of data like that in (32), seen in languages like Turkish (Inkelas 1996; Inkelas and Orgun 1995). Inkelas argues that there is necessarily a three-way contrast in voicing. Some stem-final stops show a t/d alternation (32a), with [t] appearing in codas and [d] appearing in onsets. Inkelas convincingly argues for an underlying segment that has all the features of a coronal stop, but is unspecified for [voiced]. She denotes this feature bundle as /D/. She states that the segment is assigned the value [−voiced] in codas, and [+voiced] elsewhere. Other stem-final stops consistently surface as [t] and thus are posited to be /t/ underlyingly (32b), and others surface as [d] consistently, and are thus posited to be underlying /d/ (32c).

(32) Turkish voicing alternations

a. Alternating: [Øvoiced] (unmarked for [voiced]) /D/
kanat 'wing' *kanatlar* 'wing-plural' *kanadım* 'wing-1sg.poss'

b. Non-alternating voiceless: [−voiced] /t/
sanat 'art' *sanatlar* 'art-plural' *sanatım* 'art-1sg.poss'

c. Non-alternating voiced: [+voiced] /d/
etüd 'etude' *etüdler* 'etude-plural' *etüdüm* 'etude-1sg.poss'

The point is that the two-way contrast presented to a learner of a language like Russian forces a different rule formulation than that forced by the pattern seen in Turkish. The existence of a single LAD constrains, or rather determines, the hypotheses a learner can make.

8.9.6 *Conclusions on Conciseness*

We can conclude that the Conciseness Condition is not a principle of UG. This is good, since it appears not to work (as recognized even by K&K 1979: 338):

> ...it is not conciseness per se that is involved in giving the correct formulation of a phonological rule. Rather, it is a complex and little understood set of considerations commonly referred to as rule naturalness or optimality.

A rule's conciseness or long-windedness turns out to be determined strictly by the data and the LAD, which makes use of basic set-theoretic operations. The logic of the Subset Principle requires that learners posit the most highly specified rule that is consistent with the data.

By taking seriously the idea that UG is just the LAD we no longer need to appeal to vague notions of markedness, naturalness or optimality. We just have to figure out what formal operations the LAD performs on the representations provided by the transduced data (the PLD).

Finally, it is worth reiterating that certain aspects of phonological knowledge cannot be determined by observing speakers' behavior. However, an explicit theory of the learner's initial state and an explicit theory of the learning algorithm can provide a hypothesis concerning what is not directly observable. Such hypotheses can only be formulated in the context of a theory that takes the competence/performance distinction seriously, that is, a theory of I-phonology.

8.10 Conclusions

It is useful again to make an analogy to see that characterizing UG in terms of constraints on possible grammars, instead of in positive terms, is potentially misguided. When a physicist claims that there are, say, five types of fundamental particles, s/he is not explicitly claiming that no others exist—it is impossible to know everything that exists (inductive uncertainty again). What is being claimed is that all known phenomena (within the relevant domain) can be explained using these five particle types, and so there is no reason to posit any others. Similarly, we can now propose the hypothesis that identity and non-identity conditions can be part of phonological rules, but we do not need to claim that those are the only conditions.

The philosophical arguments against constraints are bolstered by the empirical arguments given in the chapter concerning Ocp effects (developed in Reiss 2003*b*). These can be summarized as follows. The invocation of universal constraints depends upon a notion of relative ill-formedness or markedness. Such a notion cannot be justified empirically. There are rules that seem to be blocked if their output would violate the Ocp, as well as those that seems to be blocked only if their output would *not* violate the Ocp, so there is no reason to grant primacy to one type over the other. So without markedness, universal constraints are unjustified. Language-specific constraints are unnecessary, since their effects can be captured by a more precise formulation of rules.

In Section 8.9, we saw that ridding UG of constraints does not leave the child in a situation where the hypothesis space for language acquisition is unconstrained. By positing maximally restrictive rules, the child can converge on a grammar in a finite amount of time. It is worthwhile to compare the approach proposed here to that presented in an influential pre-OT paper by a phonologist who is one of the most important contributors to the success of OT. McCarthy (1988: 84), in an exposition of feature geometry, states that "The goal of phonology is the construction of a theory in which cross-linguistically common and well-established processes emerge from very simple combinations of the descriptive parameters of the model." For example, "Assimilation is a common process because it is accomplished by an elementary operation of the theory—addition of an association line" (86). After attempting to motivate two operations and two constraints on well-formedness, McCarthy declares that "each operation and constraint is predicted to operate on each class node of the feature geometry in some reasonably well-attested linguistic phenomenon" (90).[24] The vagueness of terms like *common, well-established*, and *reasonably well-attested* should alert us to the lack of rigor inherent in such an approach. A simpler, more explicit approach is to figure out what is the minimum amount of representational and computational machinery needed to generate attested patterns. Rather than seeing this as an original suggestion, it strikes me as "the natural approach: to abstract from the welter of descriptive complexity certain general principles governing computation that would allow the rules of a particular language to be given in very simple forms, with restricted variety" (Chomsky 2000: 122).

With this goal in mind, phonology should not return to the rules-and-constraints models that predate Optimality Theory, but to a pure rule-based formalism. The nature of the types of rules needed by phonological theory thus becomes an empirical question that promises to yield answers if not prejudiced by preconceived notions of what rules "should" look like.

Instead of the taxonomic generalizations offered by spurious markedness-based theories like OT, the approach advocated here will offer deeper insight into the nature of phonological computation. Such insight is the goal of cognitive science in general:

> [I]f we confine ourselves to the scientific and intellectual goals of understanding psychological phenomena [as opposed to predicting observed behavior–cr] one could certainly make a good case for the claim that there is a need to direct our attention away from superficial "data fitting" models toward deeper structural theories (Pylyshyn 1973: 48).

[24] The following sentence is much closer to a coherent proposal: "In other words, we should be able to freely combine the predicates of our theory of representations and our theory of operations and constraints and, in each case, come up with some real rule that languages have." See, however, Hale and Reiss (2000*a*, *b*) for arguments that the set of actually attested languages is expected to be only a subset of the set of computationally possible human languages allowed by UG.