

UC Santa Cruz

Wondering at the Natural Fecundity of Things: Essays in Honor of Alan Prince

Title

On Theoretical Facts and Empirical Abstractions

Permalink

<https://escholarship.org/uc/item/1066p928>

Author

Smolensky, Paul

Publication Date

2006-06-20

On Theoretical Facts and Empirical Abstractions*

Paul Smolensky
Johns Hopkins University

1. Multi-level Theory Evaluation

My goal is to argue the merits of a type of work that is somewhat rare in linguistics, and to illustrate this kind of work in three domains: phonological inventories and conjunctive constraint interaction (Section 4.1), non-participating segments in vowel harmony (Section 4.2), and the general nature of phonological categories like ‘possible word of language L ’ (Section 4.3).

The argument begins with claim (1).

(1) Theory evaluation must be carried out at multiple levels of abstraction.

I will illustrate shortly what is intended here by ‘levels of abstraction’, but the phrase ‘abstracting away from X ’ captures the basic idea rather well: at higher levels of abstraction, more X s have been abstracted away.¹

I do not take (1) to be a particularly controversial claim, but I will give two types of arguments for it, one (Section 1.2) concerning the development of the theory over time and the other (Section 5) concerning the payoffs for linguistics in connecting with several related disciplines.

Of central interest is not the argument *for* (1) but the argument *from* it. For (1) has two interesting consequences. First, theory evaluation at multiple levels of abstraction requires uncovering *the facts about the theory* at multiple levels. Evaluation involves crucial questions such as ‘does theory Y exist?’², ‘is theory Y equivalent to theory Z ?’ and the like. These are questions of *theoretical fact* that need to be asked at multiple levels of generality, ranging from the most general level — defined by the fundamental principles of the theory — all the way down to the level of detailed accounts, within the theory, of complex linguistic patterns. It is convenient to speak (loosely: note 1) of the theory at the most general level as the *0th-order theory*; at the next-most-general level, *the 1st-order theory*, and so on. The *n*th-order theory is a refinement of the (*n*–1)st-order theory, providing somewhat more detailed characterizations of somewhat less abstract theoretical entities. (N.B.: In the perspective adopted here, constructing theories of increasingly higher order means building theories that are *less and less*, not more and more, abstract.)

The second interesting consequence of (1) is that theory evaluation at multiple levels of abstraction includes (among other important things) tests of empirical adequacy at multiple levels of abstraction. This in turn requires identifying *the empiri-*

* These remarks take their inspiration and starting point from Prince 2006, although I cannot pretend to achieve a comparable degree of penetration. All errors and shortcomings are mine, and inadequacies of the present paper should not be taken to reflect on Prince’s work.

¹ The scale of levels referred to in (1) is emphatically not of the type ‘phonetics < phonology < morphology < syntax < ...’ or ‘feature < segment < mora < syllable < ...’. The term ‘level’ is in fact not ideal as it suggests a totally ordered sequence; a partial ordering, ‘refinement’, is more accurate: if theory-version T_1 ‘abstracts away from X_1 ’ while theory-version T_2 ‘abstracts away from X_2 ’, where X_2 is a proper subset of X_1 , then T_2 is a refinement of T_1 . With this understood, in the text I will continue to talk as though theory-versions were totally ordered, as in the terminology ‘*n*th-order theory’ below.

² The urgency of this particularly pointed question is brought out in Prince 2006.

cally correct characterization of linguistic phenomena at multiple levels of abstraction. These levels include the lowest level — fully detailed characterization of a particular phenomenon in a particular language — but also more general levels in which details are increasingly ‘abstracted away’ to leave increasingly abstract linguistic descriptions capturing increasingly more general empirical generalizations.

The notion of multi-level theory evaluation under consideration here is fully precise at all levels: increasingly *abstract* levels are not increasingly *vague* levels. It is the job of the n th-order theory to exactly explain the n th-order empirical generalizations: to formally derive these generalizations as theorems from the n th-order principles. The empirical generalizations that are adequate to this task are not vague statements about detailed data: they are precise statements about abstract data — statements I will call *empirical abstractions*. In order to evaluate the n th-order theory, we need to work out the facts of that theory, including its predictions, and show that they correctly characterize the n th-order empirical abstractions.³

1.1 Syllable Theory

An example of this conception of multi-level theory development is provided by the analysis of syllable structure in Prince and Smolensky 1993/2004 (henceforth P&S).⁴ Three levels of abstraction can be discerned. The most general is the level of *Basic Syllable Theory* (P&S: Chapter 6). This is the 0th-order theory, in which linguistic structures are what I will call *Basic Syllables*, structures of the form $[_\sigma(\text{Onset}) \text{Nucleus} (\text{Coda})]$.⁵ At this level, the principles are constraints like ONSET and PARSE. Thus the 0th-order syllable $[_\sigma \text{Onset Nucleus}]$ consists of an onset position (satisfying the constraint ONSET), a nucleus position, and no coda position (satisfying NOCODA). In this 0th-level syllable theory, the distinction between simplex and complex onsets has been abstracted away. The 0th-order empirical abstraction is the Jakobson Typology of basic syllable structures.

One natural next-order theory would refine the syllable structures to distinguish complex from simplex onsets, nuclei, and codas; new constraints such as $*\text{COMPLEX}_{\text{Onset}}$ would be added to the constraints of the Basic Syllable Theory. The 0th-order syllable $[_\sigma \text{Onset Nucleus}]$ would be refined to several distinct syllable

³ The n th-order theory and the n th-order empirical abstractions are themselves completely precise formal objects. The n th-order empirical abstraction will however typically be an approximate and not an exact generalization of the $(n+1)$ st-order empirical abstractions. Thus the empirical abstraction studied in Section 4.1, the harmonic completeness of phonological inventories, is an approximation to the more complex descriptions of inventories at more concrete levels. To claim that an empirical abstraction at level n is ‘correct’ is to claim that it is the most accurate approximation to the facts that is available given the distinctions that have not been abstracted away at level n .

⁴ Another example is provided by works such as McCarthy and Prince 1993, 1995 which develop highly general theories that explain prosodic morphological phenomena via principles such as the Generalized Template Hypothesis implemented in OT via general ranking schemas relating markedness and faithfulness constraints that regulate morphology, prosody, and segmental phonology.

⁵ For illustrating the notion empirical abstraction, it is best to regard the syllables of Basic Syllable Theory as a sequence of syllable positions, rather than a sequence of abstract segments ‘C’ and ‘V’.

types: [CV], [CCV], [CVV], etc. All distinctions among different consonants, and all distinctions among different vowels, would be abstracted away, as they are in the 0th-order theory. The empirical abstractions at this level would be typological generalizations concerning such syllable types, such as those summarized in Blevins 1995.

As it happens, this refinement was not pursued in P&S; there, the 1st-order theory concerns abstract segments distinguished solely by their position along a sonority scale (Chapter 8), with constraints harmonically aligning the sonority scale to the structural Peak-Margin distinction within the syllable. The 1st-order empirical abstraction in P&S is this: (i) inventories of segments permitted in different syllable positions are characterized by a minimal sonority level for nuclei and a maximal sonority level for onsets, and (ii) the set of segments allowed in coda is a subset of those allowed in onset. It is a well-defined question of theoretical fact whether the proposed sonority-based syllable theory predicts the target empirical abstraction. Showing that the answer is affirmative turns out to require considerable analysis.

Finally, the 2nd-level syllable theory of P&S concerns segments specified by phonological features including Place as well as sonority-determining features, subject to additional constraints such as Coda Conditions on Place. The 2nd-level empirical abstractions treated in P&S include syllabically-governed truncation and augmentation patterns in the nominative paradigm of Lardil at a level of concreteness typical of phonological analysis (Chapter 7).

The example of the P&S analysis of syllable structure illustrates the point, emphasized above, that even at the most abstract level, both the theory and the empirical abstractions deployed are not in the least vague, but utterly precise. The Jakobson Typology is a completely formally specified empirical abstraction, and the Basic Syllable Theory is also completely formally specified; it is a fully well-defined question to ask whether it is a fact of the theory that the typology it predicts matches the empirical abstraction. The Jakobson Typology is not a vague description of syllables built of featurally-specified segments: it is a precise characterization of the world's syllables at the level of onset, nucleus and coda positions. And the Basic Syllable Theory is not a vague description of general principles governing the fully-specified syllables of languages such as Lardil; it is a precise set of principles governing the abstraction *Basic Syllable*. Nonetheless, the concrete, 2nd-level syllable theory proposed in P&S to account for Lardil (Chapter 7) and Berber (Chapter 8) is a refinement of the 0th- and 1st-order theories, incorporating the constraints of those theories, now interpreted as applying to fully-specified segments.

The topics of theoretical facts and empirical abstractions are discussed in Sections 2 and 3 respectively. Three examples of theory evaluation at abstract levels are presented in Section 4. Speculation concerning some potentially important potential payoffs for linguistics deriving from the pursuit of multi-level theory development is the topic of Section 5. Before moving on, however, I digress to present one line of argument supporting the basic premise (1).

1.2 The Import of Multi-level Articulation of Theory and Data

It is not particular data points but rather empirical generalizations that it is the burden of theories to explain. An individual datum is of no inherent interest; it is the general patterns in data that constitute a science's empirical base. The crucial point is that in domains of any depth, these patterns exist at multiple levels of generality and

abstractness. And the theory must explain each generalization at the level of abstractness appropriate for that generalization. That is, the theory and the empirical generalizations must both be articulated at multiple levels of generality.

Indeed, patterns at higher levels of generality are arguably more important for the theory to explain: it would appear that a theory that can account for detailed patterns in the data but cannot explain the most general patterns is deeply deficient.

Explanations at higher levels of generality typically play more prominent roles in tracking the cumulativeness of scientific progress. Sorting out the story — formally explaining empirical abstractions — at higher levels of generality is more likely to be a relatively stable accomplishment than accounting for more detailed patterns, which will typically take longer to settle in a lasting fashion. Even at times of theoretical revolution in a science, the highest levels of the theory are most likely to provide continuity and guidance through the transition.⁶ And if a revolution is profound enough to shake up the theory at the highest levels of generality, it is impossible to understand the true nature of the revolution unless it is clear what the 0-th order theory was before, what it is after, and what the ramifications of this shift are down to *n*-th-order theories.⁷

The net result of such considerations would seem to be that, on both theoretical and empirical fronts, neglecting higher levels of abstraction is not a sound policy in the long run.

2. Theoretical Facts

My argument here seeks to build on Prince 2006, which forcefully argues the importance of working out what I am calling here ‘theoretical facts’. Prince calls research directed at discovering such facts the *Analytical Method* and theories for which such research is possible *Free-Standing Theories*. Prince contrasts this with the *Descriptive Method*, in which the consequences of a theory are explored solely by applying that theory to data. This is the only method available for studying what Prince calls *Theories of Data*, which are essentially generic formalisms that have virtually no consequences except when in direct contact with data. Prince observes the following curious fact: so ingrained is the Descriptive Method in the tradition of linguistic research that it has been the overwhelmingly dominant practice even in the study of Free-Standing Theories, where the Analytical Method is available.

2.1 The Analytical and Descriptive Methods in Optimality Theory

Optimality Theory is an interesting case in this respect. In P&S, the Descriptive Method was deployed to some degree, for example, to show how the OT theory of constraint interaction makes it possible to shed fresh light on even such classic prob-

⁶ Even in as dramatic a revolution as the transition from classical to quantum mechanics, many of the most general principles survived, such as the conservation of energy and momentum. The Correspondence Principle guided the development of the new theory by requiring at the level of the most general laws that the new theory give results consistent with the old theory in the limit where the microscopic approaches the macroscopic.

⁷ In the transition from classical to relativistic mechanics, there was a major shift in the status of the most general physical variables — spatial and temporal intervals — from absolute to observer-relative. Getting this high-level change sorted out properly was essential to the working out of the more detailed principles of relativistic physics.

lems as Latin stress (Chapter 4), and to cope with even aggressive alternations such as those found in the Lardil nominative (Chapter 7). In the subsequent OT literature, this style of research has been quite actively pursued.

It is worth observing that in P&S even the application of the Descriptive Method bears strong marks of the Analytical Method. How is OT deployed to describe data within a language? Particular constraints are proposed and a ranking of these constraints is shown to yield the observed forms as optimal outputs when particular underlying forms are fed as input to the ranking. How are the constraints adduced from the data? How is the ranking determined? How are the optimal outputs determined?

These three general questions are all in the domain of the Analytical Method, as their answers constitute facts of the theory. In P&S, such analysis yielded theoretical facts that were explicitly deployed in, for example, deriving the description of the Lardil pattern. No ranking can possibly yield the unfaithful mapping $I \rightarrow O$ unless there exists at least one markedness constraint preferring O to a fully faithful parse O_F ; this general fact of the theory licenses certain conclusions about the kinds of constraints that must be at work in the particular data under examination. According to the Cancellation/Domination Lemma of P&S (Chapter 8), the mapping $I \rightarrow O$ cannot occur unless, for any competing output O' , every mark of constraint violation incurred by $I \rightarrow O$ is either cancelled or dominated by a mark incurred by $I \rightarrow O'$. This theoretical fact makes possible deduction of the ranking, which is informed as well by another general fact of the theory derived in P&S, Pāṇini's Theorem on Constraint Ranking (Chapter 5). Finally, given a proposed ranking of proposed constraints, determining whether the desired form is indeed the optimal output for a given input is made possible by the Method of Mark Eliminability (Chapter 7), a fact of the theory which, unlike enumeration of a few plausible competitors in a tableau, actually enables determination of optimality.

Correct deployment of a theory for data description is only possible if theoretical facts such as those just mentioned are systematically and carefully exploited. And this is of course only possible if such facts of the theory have been brought to light, which requires vigorous exercise of the Analytical Method. Thus it is perhaps remarkable that in contrast to the Descriptive Method, the Analytical Method has not been widely practiced in the OT literature following the work in P&S that uncovered the Cancellation/Domination Lemma, Pāṇini's Theorem, the Method of Mark Eliminability, and other general facts of Optimality Theory.

There are some notable exceptions, the most systematic being the work of Prince and collaborators. Questions of the relation of the special to the general, as addressed by work on Pāṇini's Theorem begun in P&S, are pursued extensively in Prince (2001) and preceding work. The important high-level explanatorily-driven constraints imposed by the structure of OT on the constraints in universal grammar are presented in Prince 1997. The character of optima and the logic of learning rankings given a set of optima is analyzed by Prince and collaborators to a depth rarely seen in research using the Analytical Method in linguistics (Samek-Lodovici and Prince 1999; Prince 2002b, 2002c; Samek-Lodovici and Prince 2002). This ranking logic enables work within the Analytical Method on the problem of algorithms for learning OT grammars (Tesar et al. 2003; Tesar and Prince 2003; Prince and Tesar 2004). Taken together, this body of research reveals with unprecedented clarity a central set of general facts of a linguistic theory, theoretical facts which, like those

uncovered in P&S, have dual empirical relevance: they provide the tools for correct deployment of the theory in describing detailed empirical patterns, and they characterize the shape of the empirical abstractions predicted by the theory at a highly general level (see, e.g., the Subordination Spectrum of Prince and Smolensky 2003).

With respect to theory-relation questions (such as ‘is theory *Y* equivalent to theory *Z*’), Prince’s work has also shed much light on the connection between optimization under strict domination in OT and numerical optimization such as that performed by neural networks (Prince and Smolensky 1997; Prince 2002a, 2005) as well as the relation between connectionist (Rumelhart and McClelland 1986) and generative theories of the character and acquisition of morphophonological knowledge (Pinker and Prince 1988; Prince and Pinker 1988a; Prince and Pinker 1988b). Prince’s work has also analyzed in great depth a neural network model of stress and syllabification (Goldsmith and Larson 1990; Goldsmith 1992, 1993) and identified many remarkable general theoretical facts of this model (Prince 1993, 1996, 2006).

2.2 The Analytical Method and Factorial Typology

Despite the major progress discussed in the previous section, there remains urgent need for further deployment of the Analytical Method in OT. One key example concerns Factorial Typology. A toolkit of general methods for deducing the Factorial Typology predicted by a given substantive OT proposal is badly needed. Empirical evaluation of such a proposal rests precisely on the extent to which the predicted Factorial Typology matches the corresponding empirical abstraction. Yet in many important cases, computing the Factorial Typology is simply beyond the reach of the Analytical Method given the current knowledge of the facts of OT. Presumably, some of the techniques in such a toolkit would constitute calculational procedures the consequences of which can only be determined in the context of particular data; others would presumably characterize the general properties of the Factorial Typology predicted by any set of constraints possessing some general properties identified by the Analytical Method.

Tools of the latter kind are needed to cope with situations such as this. In a course in OT syntax, Géraldine Legendre set her students the task of using OTSoft (Hayes, Tesar, and Zuraw 2003) to find the combined typology predicted by the proposal in Vikner 2001 for *do* support, subject-auxiliary inversion, and negation. This exercise produced widespread panic when it was discovered that the predicted number of languages varied widely from student to student, ranging as high as 430. (This arose because the students provided different candidate sets as input to OTSoft.) What was the right answer, they were desperate to know. A systematic listing of all inputs produced a prediction of 164 languages, a result that did little to calm the students’ panic. But this result reflected the fact (pointed out by OTSoft in fact) that quite a few candidates had the same constraint violation profile, so in many cases two languages in the typology differed only in the selection of which harmonically-equivalent candidate was identified as ‘the’ winner. Factoring out this component of the typology is already an application of a theoretical fact, albeit an extremely basic one: no ranking will distinguish two candidates with the same Harmony (constraint violation profile). Rather than returning the full set of (harmonically equivalent) optimal outputs for a given input, OTSoft was selecting individual outputs from these optimal sets and treating different choices as different predicted languages.

Reconstructing the input to OTSoft to eliminate every output that is harmonically equivalent to an already-listed candidate reduced the predicted number of languages to 18. This was still a somewhat disturbing result to the students, perhaps because their understanding of Vikner's analysis suggested there were fewer degrees of freedom than this number seemed to indicate. The correctness of this suspicion is revealed by showing that the number 18 reflects free combination of a small number of possibilities along a few independent dimensions. In fact, the typology can be reduced to 5 core languages, with two additional independent dimensions each with 2 variants; for the 4 core languages in which both these additional dimensions are relevant, this yields $4 \times 2 \times 2 = 16$ variants; for the 5th core language, only one of the dimensions is relevant, yielding an additional $1 \times 2 = 2$ languages, for a total of 18. (One dimension of variation is whether subjects move, the other is whether verbs move.)

A useful general analytic tool would be one enabling the factorization of the Vikner typology — a tool showing how the typology can be understood as a very small set of core languages, each with variants differing along independent dimensions. This factorization would presumably follow from showing that the Vikner analysis exhibits a general configuration of constraint violations amounting to independence of the consequences of relative rankings of disjoint subsets of constraints. (The Jakobson Typology of Basic Syllable Structures exhibits such a factorization; the relative constraint rankings determining whether onsets are required are independent of the constraint rankings determining whether codas are forbidden, so these two dimensions vary independently within the typology: P&S Chapter 6.)

3. Empirical Abstractions

Intimately connected with the contrast between the Analytical Method and the Descriptive Method is the nature of the data being explained. The Analytical Method yields formal results: facts of the theory. The empirical adequacy of the theory is assessed by comparing these facts to data; definitive comparison requires that the data be as sharp as the theoretical facts.

The key issue is whether an empirical generalization is a vague statement about concrete facts or an exact statement about abstract facts. If the former, then the clear route is to account for the generalization by accounting for concrete facts — this is the Descriptive Method. If, however, an empirical generalization is an exact statement about abstract facts — what I'm calling an empirical abstraction — then to explain it is to derive it formally from precise principles of the theory; that is, to identify the empirical generalization as a fact of the theory deduced by the Analytical Method.

Crucial to the formulation of empirical abstractions are abstractions of the *objects* of empirical investigation. A few illustrations of this point from P&S were given in Section 1; the 'C' and 'V' of CV syllable theory are such abstract objects, and the Jakobson Typology of Basic Syllable Structures can only be stated once these abstract objects are available.

The potential long-term importance for linguistics of empirical abstractions and the abstract objects figuring in them may be suggested by the importance of such abstractions for theoretical physics. While it is a painfully over-used (and much abused) cliché to cite physics as the model of a highly developed discipline both

theoretically and empirically, I think it is rarely appreciated that this state of development is made possible only by empirical abstractions. So I digress to illustrate.

It is no exaggeration to assert that the large majority of the material studied in undergraduate physics training, and a substantial portion of the material studied in graduate training, concerns the identification of general facts about physical theories and the comparison of these facts to empirical abstractions at some considerable distance from the concrete. Here are a few examples from elementary mechanics.

Newton's Laws of motion and of gravitation concern the abstraction *point mass*. Like a point in Euclidean geometry, such an abstraction has zero extension in all directions; nonetheless it has non-zero, fixed mass.⁸ *Free fall* is an abstraction in which a point mass moves under the influence of a uniform force field ("gravity") only. The *simple pendulum* is an abstraction in which a point mass is subject to a uniform force field ("gravity") and a center-directed force ("of a taut string") that limits its motion to the arc of a circle of some radius ("the length of the string"). The uniform force field modeling gravity both in free fall and in the simple pendulum is an approximation to the gravitational force field actually determined by Newton's Law of Gravitation, a force field that is not uniform, but rather radially directed to the center of a point mass ("the earth") generating the field; if the pendulum moves through a region of space small in comparison to the distance of this region from the center of gravitational force, the field is uniform, to a high degree of approximation.

An abstraction studied in great depth in a wide range of formalisms is the *harmonic oscillator*, a point mass confined to move on a (zero-width) line and subject to a force directed toward an equilibrium point such that the magnitude of the force grows in proportion to the distance of the mass from the equilibrium point. This is usually first introduced as an approximate description of the force exerted by a spring that is compressed or stretched from its resting length; in reality, this type of force is the first-order (linear) approximation to a very general class of smooth force fields around an equilibrium point, and this is the basis of the extreme generality of the system. (To take one example far removed from springs, the quantum theory of the electromagnetic field — e.g., light — is based on harmonic oscillators.)

The gravitational field is an instance of a large class of abstractions called *conservative systems* in which the force on a point mass is the negative gradient of a function V : at any point x , the component of the force along any axis is minus the rate of change of V at x along that axis). For these systems, it can be proved from Newton's Laws of Motion that the total energy of a point mass m is unchanged as it moves; the total energy is an abstraction defined in this context as $E = \frac{1}{2}mv^2 + V$ (where v is the speed of the point mass). This is a simple instance of the Law of Conservation of Energy; it holds for a great range of abstract systems, once other types of energy are added to E (thermal energy, the energy of the electromagnetic field, etc.). In the abstraction of *elastic collisions*, colliding point masses rebound in a particular way — a way defined so that energy is conserved.

The point is that the core of theoretical physics — which by now fills dozens of courses — is entirely directed to empirical abstractions. The physics of real material bodies moving in real air under real gravity exposed to real electromagnetic fields is

⁸ Thus the density — mass per unit volume — of every point mass is infinite, but point masses nonetheless differ in their masses.

only reached after layer upon layer upon layer of abstract systems have been thoroughly understood by aggressive pursuit of the Analytical Method. The physics of real systems is built entirely upon a great many layers of abstract entities and the principles governing them. Without such a multi-layered wealth of abstractions, theoretical and empirical, a deeply developed physics is unimaginable. That physics depends on idealizations and approximations is typically recognized, but the concomitant commitment to many levels of empirical abstractions appears to be less widely appreciated.

4. Abstract Theory Evaluation: Examples

In this section, I will briefly describe three examples of empirical abstractions in phonology, and the theoretical facts that I have proposed to explain them. These examples concern linguistic inventories (Section 4.1), linguistic “processes” (Section 4.2), and linguistic categories (Section 4.3).

4.1 Basic Inventories: Strong Harmonic Completeness and Local Conjunction

A natural 0th-order abstraction for an inventory of linguistic elements is the set of all n -tuples of possible values for n binary features $\{\phi_i\}_{i=1, \dots, n}$: this set is \mathbf{B}^n , where $\mathbf{B} \equiv \{+, -\}$. (There are numerous obvious extensions and refinements, including non-binary features, unspecified features, incompatible feature values, etc.) The ‘features’ can be imagined to be standard segmental features ($[\pm\text{back}]$) or abstract ‘features’ describing structural options (such as $[\pm\text{long}]$ or $[\pm\text{extra-syllabic}]$ or $[\pm\text{fronted-subject}]$).

In OT, the essential core of constraints evaluating such objects are the faithfulness constraints $F_i \equiv \text{IDENT}[\phi_i]$ and the markedness constraints $M_i \equiv *[\phi_i]$ (where, without loss of generality, we may assume all features to be defined so that the marked value is $+$). (Again, there are many obvious extensions and refinements, such as feature-co-occurrence constraints, MAX/DEP constraints, etc.)

A *Basic OT Typology* is defined to be an OT system with the core constraints $\{F_i\}$, $\{M_i\}$ evaluating $I \rightarrow O$ mappings defined by all pairs (I, O) with I and O any element of \mathbf{B}^n . A *Basic OT Inventory* is a set of elements I in \mathbf{B}^n such that, for some ranking of the core constraints, $x \in I$ iff and only if $x \rightarrow x$ (for the input x , the optimal output is x itself; under Richness of the Base, when y is not in the inventory, we have $y \rightarrow z$ for some $z \neq y$). These definitions are summarized in (2).

(2) *0th-order theory: Basic OT Inventory Typology*

- a. $Con \equiv \{F_i\}_{i=1, \dots, n} \cup \{M_i\}_{i=1, \dots, n}$
- b. A *Basic OT Inventory* is a set $I_{\mathcal{H}} \subset \mathbf{B}^n$ defined by $I_{\mathcal{H}} \equiv \{x \in \mathbf{B}^n \mid x \rightarrow_{\mathcal{H}} x\}$ where \mathcal{H} is a ranking of Con and ‘ $\rightarrow_{\mathcal{H}}$ ’ denotes the mapping $\mathbf{B}^n \rightarrow \mathbf{B}^n$ defined by optimization over \mathcal{H} .
- c. The *Basic OT Typology* is the set of all inventories $\mathcal{T} \equiv \{I_{\mathcal{H}} \mid \mathcal{H} \text{ is a ranking of } Con\}$.

The two empirical abstractions I propose to consider are given in (3). Both are highly general expressions of the Prague school finding that inventories are shaped by markedness: if a more marked element x is present in an inventory, so is a less marked element y . That is, inventories are *harmonically complete*. (For a recent case

study, the harmonic completeness of the rhymes of English monomorphemic monosyllabic words, see McClelland and Vander Wyk 2006.)

(3) *Empirical abstractions*: Basic Inventories

- a. A Basic Inventory is Harmonically Complete.
- b. Basic Typologies are Strongly Harmonically Complete (SHARC).

The generalizations in (3) derive directly from P&S: Chapter 9. The relevant definitions are given in (4). Consider first (3a). In a harmonically complete inventory I , when $y \succ x$, if x is present in I , then so must be y . ‘ $y \succ x$ ’ means ‘ y is more unmarked than x ’; markedness constraints prefer y to x . Precisely, this means that (i) for some markedness constraint, y is preferred to x , and (ii) for no markedness constraint is x preferred to y . In a Basic OT Typology, this means that (i) on some markedness constraint M_k , y is preferred to x , i.e., the value of ϕ_k for x , written x_k , is the marked value $+$ while the corresponding value y_k is the unmarked value $-$, and (ii) for every feature ϕ_i for which y has the marked value ($y_i = +$), the corresponding value of x is also the marked value ($x_i = +$). This is spelled out in (4a).

(4) *Definitions*. Harmonic completeness

- a. A Basic Inventory I is *harmonically complete* if and only if
 - i. $x \in I$ and $y \succ x$ implies that $y \in I$, where
 - ii. $y \succ x$ if and only if for each ϕ_i such that $y_i = +$, $x_i = +$, and for some ϕ_k , $x_k = +$ and $y_k = -$.
- b. A Basic Typology \mathcal{T} is *Strongly Harmonically Complete (SHARC)* if and only if

for any subset $I \subset \mathbb{B}^n$, $I \in \mathcal{T}$ if and only if I is harmonically complete.

The empirical abstraction (3a) states that every Basic Inventory is harmonically complete. What of the converse? Is every harmonically complete inventory a Basic Inventory, that is, present in the Basic Typology? If so, then the Basic Typology has the *SHARC* property: it is *Strongly Harmonically Complete* (4b). If inventories are indeed determined by markedness, we expect the SHARC property to hold: this is just what is asserted in the empirical abstraction (3b).

Are the empirical abstractions (3) in fact explained by Basic OT Typology theory? Calling upon the Analytical Method to analyze this Free-Standing Theory, we find the theoretical facts in (5).

(5) *Theoretical facts*: Basic OT Typologies

- a. Every Basic OT Inventory is Harmonically Complete.
- b. Basic OT Typologies are *not* Strongly Harmonically Complete.

The analysis deriving (5) is presented in Smolensky 2006. While (5a) might be expected, (5b) may be surprising. The problem for Basic OT Typologies comes from inventories that ban only the worst of the worst: *BOWOW inventories*. A simple example is the fragment $I_0 \equiv \{t, s, k\}$ of the English obstruent inventory, where the features are $\phi_1 \equiv [\text{continuant}]$, $\phi_2 \equiv [\text{velar}]$. The problem is that any ranking of the constraints $\{M_1, M_2, F_1, F_2\}$ which admits the marked segments s and k into the inventory also admits x (the velar fricative) as well. x is ‘the worst of the worst’ in that it is the segment that is most marked with respect to $[\text{velar}]$ of the set of segments that are

most marked with respect to [continuant]: $\{s, x\}$. The English inventory bans only the worst of the worst. This BOWOW inventory cannot be generated by any ranking of the core constraints defining the Basic OT Typology \mathcal{T} : I_0 is not part of \mathcal{T} . But I_0 is harmonically complete: I_0 contains the ϕ_1 -marked segment s so harmonic completeness requires that I_0 also contain the corresponding ϕ_1 -unmarked segment t , which it does; I_0 contains the ϕ_2 -marked segment k so harmonic completeness requires I_0 to contain the corresponding ϕ_2 -unmarked segment t , which it does. Since I_0 is harmonically complete, Strong Harmonic Completeness requires that it be in the typology \mathcal{T} , but it is not: \mathcal{T} is not SHARC.

A simple extension of the Basic OT Typology, however, is SHARC. If the constraint set includes not only the core constraints, but also *local conjunctions* of markedness constraints, we get the *Basic OT-LC Typology*. For the fragment of the English obstruent inventory I_0 , the local conjunction arising from the two markedness constraints is $C \equiv M_1 \& M_2 = *[\text{+continuant}] \& *[\text{+velar}]$. This constraint is violated whenever its conjuncts are both violated in the same segment; of course x violates C but t , s , and k do not. Adding C at the top of any ranking of the core constraints that admits $\{t, s, k\}$ into the inventory yields a ranking that excludes x , i.e., a ranking that generates I_0 . Indeed the general analysis of Smolensky 2006 derives the general result in (6).

(6) *Theoretical fact*: Basic OT-LC Typologies

A Basic OT-LC Typology is Strongly Harmonically Complete.

Thus the empirical abstractions of (3) figure directly in an argument concerning the fundamental structure of OT, the argument of Smolensky 2006 that constraint interaction in OT must include local conjunction as well as domination.

4.2 The Harmony/Inventory Theorem

In vowel harmony, a basic observation is that some vowels participate in the harmony process while others do not. The empirical abstraction in (7) is an appropriate 0th-order generalization characterizing the set of non-participating vowels.

(7) *Empirical abstraction*: Non-harmonizing vowels

- a. A vowel can fail to participate in harmony only if doing so would take it out of the language's inventory of underlying vowels.
- b. A vowel v is in the inventory of underlying vowels of a language if and only if underlying v surfaces faithfully in the language (definition).

Can this empirical abstraction be explained by a general OT theory of vowel harmony? That the answer is affirmative is shown by the analysis of Smolensky 2006, which presents a 0th-order OT theory of vowel harmony (driven by a markedness constraint requiring ϕ -agreement of vowels in some domain, for a vowel feature ϕ) and derives a fact of this theory, (8).

(8) *Theoretical fact*: Harmony/Inventory Theorem

Let $A_{+\phi}$ denote a $[\text{+}\phi]$ vowel harmony constraint. Suppose that a language L has an $A_{+\phi}$ -process in an environment E and that some $[-\phi]$ segment s_- fails to undergo this process. Then, in E , s_+ is not in L 's underlying inventory.

The formal definitions presumed in this theorem are given in (9).

- (9) *Definitions*: 0th-order abstract empirical objects for Harmony/Inventory Theory
- a. A segment x is in the *underlying E-inventory* if in environment E , $x \rightarrow x$.
 - b. Let A be a markedness constraint. A segment x *undergoes an A-process* if, when A is applicable in environment E ,
 - i. x violates A ,
 - ii. $x \rightarrow y$, where y violates A to a lesser degree than does x , and
 - iii. y has lower Harmony than x with respect to the hierarchy of markedness constraints other than A .
 - c. A language *has an A-process* if some segment undergoes the A -process.
 - d. Let φ be a vowel feature. A $[+\varphi]$ *vowel harmony constraint*, when it applies, requires all vowels in some domain \mathcal{D} to bear the feature $[+\varphi]$. (In the standard case, such a constraint will apply when $[+\varphi]$ is borne by some controlling vowel, typically at one edge of \mathcal{D} .) Let such a constraint be denoted $A_{+\varphi}$.
 - e. A language *has $[+\varphi]$ vowel harmony* if it has an $A_{+\varphi}$ -process.
 - f. A *minimal φ -pair* are two segments x_- and x_+ that are featurally identical except that x_- is $[-\varphi]$ and x_+ is $[+\varphi]$. The notation x_-/x_+ always denotes a minimal φ -pair.

These definitions are straightforward except perhaps for (9b.iii). The point is simply that if $x \rightarrow y$ and y has lower markedness-Harmony than x even without considering the constraint A , then there is no reason to believe it is A that is responsible for this mapping of x to y ; just as possibly, it is merely a coincidence that A also favors y over x , while it is in fact some other, higher-ranked constraint that renders this unfaithful mapping optimal.

This example illustrates a general point about empirical abstractions made in Section 3: in order to obtain a formal explanation of the empirical abstraction (7) concerning non-harmonizing vowels, it is necessary to develop quite a few abstractions of linguistic structure — (9) gives a half-dozen of them. Clearly, a number of the details involved in a concrete analysis of “real” vowel harmony in a “real” language have been abstracted away. As emphasized in Section 3, this is essential to the explanatory enterprise: an empirical abstraction is not a vague generalization concerning relatively concrete linguistic objects, but a precisely formulated generalization concerning relatively abstract linguistic objects.

4.3 Basic Combinatorial Induction and Lexical Similarity Theory

The final example addresses a question at a still higher level of generality and abstraction: what is the nature of grammatical categories, and are they of the same kind as other cognitive categories? (See Pinker and Prince 1996 for an extended discussion of this question in the context of the categories *regular* and *irregular verb*.) The intuition evaluated is that grammatical categories are inherently *combinatorial*, in contrast to the *similarity*-based character of many other category systems in cognition.

This intuition, seemingly fundamental to the generative approach to grammar, has been challenged, at least implicitly, in recent exemplar-based proposals that knowledge of phonology is exhausted by knowledge of the lexicon: that putatively grammatical phenomena such as word-likeness judgments reflect not cognitive generalizations, such as phonotactics, of the sort defining grammars, but rather judgment of the similarity of a candidate lexical item to the set of actual lexical items.

Prior to the question of whether the grammatical or the lexical view is more adequate empirically, there is a basic question, like Prince's "does the theory exist?": are these views in fact *different*? — does the question of which one is correct actually exist? I suspect that part of the attraction of the lexicalist view is the intuition that the two views are not as different as grammatical theorists would like to believe: that much of the work done by grammatical computation can be done equally well — in fact, better — by computation of similarity to the lexicon. The notion "similarity" is after all highly general and flexible; surely there is *some* way of computing "similarity to the lexicon" that can explain word-likeness judgments?

This is a highly general question and it demands a highly general answer. Here is one.

I will take the following 0th-order empirical abstraction as a working hypothesis; we will return to reconsider it in light of the theoretical results.

(10) *Empirical abstraction*: Basic Combinatorial Induction (CV form)

Suppose a learner is presented with a set of CV syllables as the lexicon of a language; suppose there is no statistically significant evidence for mutual restrictions between the C and the V of a lexical item. Then when asked to judge a novel CV form, the learner will assign a relatively higher word-likeness rating to a form in which the C and the V are each separately attested in some lexical item than to a form in which either the C or the V appears in no lexical item.

Thus, for example, if the lexicon presented is $\{ti, ka\}$ then the word-likeness ratings the learner will assign to ta and ki are higher than those assigned to, say, pa or tu .

Now for a general characterization of a lexicalist view; $Lex(L)$ is the lexicon of a language L .

(11) *0th-level theory 1*: Lexical Similarity Theory

- a. All phonological knowledge particular to L is encoded in $Lex(L)$.
- b. The word-likeness of a novel item x (as a possible word of L) — $\omega_L(x)$ — is determined by the similarity of x to the individual items in $Lex(L)$.
- c. The similarity metric s is independent of L and symmetric: $s(x, y) = s(y, x)$.
- d. $\omega_L(x) = f(\{s(x, w) \mid w \in Lex(L)\})$
- e. f is a monotonically non-decreasing function:

$$s_i \geq s'_i, i = 1, \dots, N \Rightarrow f(\{s_i\}_{i=1, \dots, N}) \geq f(\{s'_i\}_{i=1, \dots, N})$$

The requirement (11c) that the similarity metric s be independent of the language L is intended to prevent s from incorporating a grammar of L (e.g., " $s(x, w)$ is high iff x and w are both generated by a particular grammar G "); incorporating a grammar of L into s clearly guts the lexicalist hypothesis of any possible interest and

does not pursue the intuition that similarity to the lexicon can replace grammatical computation. The definition (11d) states that similarity is assessed between the target form x and each lexical item w separately; all these individual comparisons are aggregated into a single word-likeness rating by the function f . And the monotonicity condition on f (11e) simply states that if the individual similarity values to the N lexical items $\{w_i\}_{i=1, \dots, N}$ of a target form x is the multi-set $\{s_1, \dots, s_N\}$ while the corresponding values for target form x' are $\{s'_1, \dots, s'_N\}$, and for each i , $s_i \geq s'_i$, then x is at least as similar as x' to the lexicon as a whole.

Again, we will return to reconsider this working definition in light of the theoretical results.

In contrast to the lexicalist view, a simple version of the grammatical view is (12).

(12) *0th-level theory 2: Basic Combinatorial Theory (CV form)*

- a. The possible phonological words of a language L form the set $\text{PWd}(L) \equiv ON \equiv \{cv \mid c \in O, v \in N\}$ where $O \subset C$ is a set of ‘consonants’, and $N \subset V$ is a set of ‘vowels’.
- b. The set of possible words of the language L determined by the primary data $\text{Lex}(L) \subset CV$ is generated by the minimal sets O, N s.t. $\text{Lex}(L) \subset \text{PWd}(L)$.
- c. If $x \in \text{PWd}(L)$ and $x' \notin \text{PWd}(L)$, then the word-likeness values ω_L satisfy

$$\omega_L(x) > \omega_L(x').$$

Applying the Analytical Method to the two Free-Standing Theories (11), (12) gives (13).

(13) *Theoretical facts: Basic Combinatorial Induction Theorem*

- a. Lexical Similarity Theory cannot explain Basic Combinatorial Induction.
- b. Basic Combinatorial Theory explains Basic Combinatorial Induction.

The success of Basic Combinatorial Theory (13b) is clear: the sets O and N of (12a) are respectively the sets of all observed Cs and Vs in the lexicon; the set of possible words PWd (12b) is thus exactly the set of all CV forms in which the C and the V have individually been observed in some lexical item. Then (12c) ensures that such forms have higher word-likeness values than all other CV forms, exactly the requirement of Basic Combinatorial Induction (10).

The success of Basic Combinatorial Theory is interesting only in that it shows that Basic Combinatorial Induction is trivial to achieve on a grammatical view of phonological knowledge; it is this contrast that makes the failure of Lexical Similarity Theory (13a) striking. The contrast also demonstrates that Lexical Similarity Theory is indeed incapable of simulating even the most basic behavior of a grammatical theory; the two theories are certainly not equivalent, nor does Lexical Similarity Theory subsume Basic Combinatorial Theory in its capacity to generate word-likeness judgments.

The proof of (13a) is somewhat involved; it is relegated to the Appendix.

With the results (13) now in hand, we can reconsider the various working assumptions. First, the theoretical assumptions. The restriction to CV forms is for con-

venience and is in no way necessary. A lexicon $\{ti, ka, pu, pu\}$ could be replaced by $\{tistem, katem, pusem, pusem\}$ or analogous forms for any prosodic shape and a directly analogous proof would follow. As shown in the proof in the Appendix, the case of four-word lexicons suffice, but if desired, examples with more than four lexical items could similarly be generated; the result cannot be escaped by insisting on larger lexicons no more than it can be escaped by insisting on larger words. The structure of the argument is very general and more general formulations of the theorem are possible; the minimal CV-form theorem and four-word lexicon proof just make the argument clearer.

Generalizing to richer notions of grammar is obviously possible; various types of phonotactic restrictions could be incorporated into the analysis. But I see no argument to the effect that the weakness of Lexical Similarity Theory relative to grammatical alternatives will be *reduced* by enriching the grammatical theory.

It may be possible to coherently characterize a version of Lexical Similarity Theory in which the similarity function s is language-particular, but in a lexicon-determined way that conceptually contrasts with a grammatical view. Note, however, that to save Lexical Similarity Theory from the negative result of (13a), such an account would need to explain how in the extremely simple cases of lexicons like $\{ti, ka, pu, pu\}$, s adopts some language-particular character that somehow invalidates the proof.

Finally, we may reconsider the empirical abstraction (10). Perhaps this is not the correct 0th-order generalization; perhaps human learners do not perform Basic Combinatorial Induction. Indeed, since they suggest that phonotactic ill-formedness in part of a word can be compensated for by well-formedness elsewhere in the word, the empirical results of Coleman and Pierrehumbert 1997 cast some doubt on the fundamental premise of Basic Combinatorial Induction. Note, however, that the proof of (13a) shows that Lexical Similarity Theory predicts on very general grounds that high well-formedness ratings can arise for forms with segments that do not appear in the lexicon, quite a striking prediction. It is no fine-grained empirical deviation from Basic Combinatorial Induction that can save Lexical Similarity Theory; gross deviations of this sort are required — deviations dramatic enough to show up even at the very high level of abstractness at which this analysis applies.

A final remark concerning the role of such abstract analysis. The work presented in this section was inspired by research of Adam Albright which aims to develop and experimentally test formal models of phonological induction from the lexicon. His preliminary empirical results suggest — contrary to the interpretation of the Coleman and Pierrehumbert 1997 results mentioned above — that exemplar-based theories of English word-likeness judgments systematically fail to capture a feature of the actual judgments: participants penalized more than exemplar theory predicts forms that violate English phonotactics when other parts of the form have relatively high similarity to a number of lexical items. This observation was quite subtle, however, and difficult to quantitatively substantiate given the enormous complexity of the actual English lexicon. The true differences of definition and prediction between Lexical Similarity Theory and grammatical alternatives, if any, are extremely difficult to identify and evaluate at the level of full concreteness of actual English. A satisfactory understanding at the concrete level can, I believe, eventually be attained, but

only if this understanding is constructed at many levels of abstraction, including highly general levels such as the one addressed in the work discussed in this section.

5. Additional Payoffs

To offer another kind of motivation for multi-level theory development in linguistics, in this final section I argue that such work can better enable linguistics to address several challenges that currently limit its ability to contribute to, and benefit from, its neighboring disciplines. I certainly do not intend to suggest, however, that linguistics-internal considerations are in any way inadequate or subordinate to linguistics-external considerations, or that the kind of linguistics-internal motivations discussed above are in need of buttressing with external justification. I consider the benefits identified below to be merely free bonuses for doing linguistics right.

There are reasons to believe that multi-level theory development can help overcome each of four significant gaps separating theoretical linguistics from its neighbors: language engineering, experimental behavioral science, computational neuroscience, and mathematical linguistics.

The first of these gaps separates theoretical linguistics from language engineering. Clear cases in which theoretical linguistics makes a crucial contribution to the success of automatic language processing systems enhance the credibility of linguistics but are quite rare. As pointed out in Prince 2006, when it comes to Theories of Data, sophisticated empiricist data-fitting techniques are hard to beat, and such techniques have been brought to very high levels indeed in contemporary statistical computational linguistics. How could linguistic theory make an essential contribution to the success of such an enterprise? Statistical learning procedures estimate parameters in probabilistic models for generating observed data; their success is thus due jointly to the strength of the parameter estimation algorithms and to the quality of the underlying generative model. There is certainly much room for major advances in the latter, as the models of language incorporated into probabilistic language engineering systems to date are by and large extremely crude. If theoretical linguistics can provide considerably better models that nonetheless admit powerful parameter estimation procedures, major improvements are to be expected. This requires finding sound models of language structure at a high level of abstraction and generality; the statistical learning procedures fill in the details. Thus the optimal level for contact between linguistic theory and language engineering will likely be a rather abstract one.

A second gap separates theoretical linguistics from experimental behavioral science. The relation between theoretical linguistics and psycholinguistics is quite different from the relation between theoretical physics and experimental physics. Coping with the subtleties intervening between linguistic competence and observable behavioral variables requires development of a complex theory of cognitive processing. Psycholinguistics has necessarily focused much of its attention on the processes constituting performance rather than the knowledge constituting competence; an understanding of such processes is interesting and important in its own right, as well as a co-requisite for experimental investigation of linguistic knowledge. As a result, experimental investigation of the competence questions at the center of linguistic theory remains largely unexplored. While the traditional empirical base for competence theory ('internal evidence') has great strengths, it also has some limitations which may ultimately prove complementary to those of experimental data. In pho-

nology, for instance, typological data on sound patterns provides an enormously rich empirical base, but, under the psychological interpretation of linguistic theory, this data bears only indirectly on the real object of linguistics: the mental systems in which phonological knowledge is manifest. For example, gaps in sound systems may be the result of historical accident, and not a reflection of mentally real phonological principles. Experimental methods may ultimately provide the means to identify which gaps are due to genuine phonological principles, in which case the data resulting from such methods would provide crucial evidence for phonological theory. Experimental methods may also enable more direct evidence concerning highly general grammatical patterns. To extend a point raised in the previous section, the correctness of Basic Combinatorial Induction may in the end be better determined through experimental tests of induction from simple ‘artificial lexicons’ than through the enormous complexities of induction from the actual English lexicon. Thus a promising arena in which to attempt to close the gap between theoretical linguistics and experimental behavioral science may be that of fundamental empirical abstractions at high levels of generality. Indeed, it would seem that experimentally grounding phonological theory at its more general levels is a prerequisite to experimentally evaluating the intricacies of alternative theories of feature geometry.

A third gap separates theoretical linguistics from computational neuroscience. It has been argued at some length that this gap is to be first bridged at a high level of abstraction (P&S: Chapter 10; Prince and Smolensky 1997; Prince 2002a, 2005; Smolensky and Legendre 2006). It is at such a level that computation in certain neural networks can be characterized as optimal satisfaction of parallel conflicting violable constraints of differing strength — connecting, obviously, with the OT characterization of grammar at a high level of generality. Indeed, like physics (considered in Section 3), computational neuroscience is a field in which a strong case can be made for the necessity and power of theory development at multiple levels, including, crucially, highly abstract levels.

Finally, a fourth gap separates theoretical linguistics from mathematical linguistics. The formal developments of mathematical linguistics have on the whole been directed to highly general questions such as the computational complexity of various grammar formalisms. Here, the important contribution of multi-level theory development involves formally precise accounts of linguistic questions at *less abstract* levels — developing accounts that extend the formal explicitness and precision of mathematical linguistics to the full range of empirical generalizations that theoretical linguistics seeks to explain. It is precisely the value of such research that has motivated this paper.

Appendix: Proof of the Basic Combinatorial Induction Theorem

The idea of the proof of (13a) is this. Consider the lexicon $Lex(L) = \{ti, ka, pu, pu\}$ (this lexicon has two homophonous words). Basic Combinatorial Induction (10) requires that the word-likeness of ta in L exceed that of bu . But bu is very similar to two words (pu, pu), while ta has a lower degree of similarity to the two closest words (ti, ka). This will imply in Lexical Similarity Theory that the well-formedness of bu exceeds that of ta , contradicting Basic Combinatorial Induction.

Proof. Define X to be the set of CV pairs which agree in either C or V, but not both:

$$(14) \mathcal{X} \equiv \{(cv, c'v) \mid v \in V; c, c' \in C; c \neq c'\} \cup \{(cv, cv') \mid c \in C; v, v' \in V; v \neq v'\}$$

Now let (x, x') be a pair in \mathcal{X} with *maximal* similarity. Let us assume this is a pair agreeing in V ; otherwise, in the rest of the proof simply exchange V s and C s. Thus $(x, x') = (c_0v_0, c_0'v_0)$ for some $v_0 \in V$; $c_0, c_0' \in C$; $c_0 \neq c_0'$. For ease of exposition, denote c_0 by b , c_0' by p and v_0 by u . Thus “ (bu, pu) ” denotes a pair in \mathcal{X} with maximal similarity:

$$(15) s(bu, pu) \geq s(cv, c'v) \text{ and } s(bu, pu) \geq s(cv, cv') \text{ for all } c, c', v, v' \text{ as in (14)}$$

Define \mathcal{Y} to be the following set of CV pairs:

$$(16) \mathcal{Y} \equiv \{(cv, pu) \mid c \neq p, v \neq u\} \cup \{(c'v, bu) \mid c' \neq b, v \neq u\}$$

Now let (y, y') be a pair in \mathcal{Y} with *minimal* similarity. y' is either pu or bu ; let's assume $y' = pu$; otherwise, in the rest of the proof simply exchange p and b . Now $(y, y') = (c_1v_1, pu)$ for some $c_1 \neq p$, $v_1 \neq u$. For ease of exposition, denote c_1 by t and v_1 by a . Thus “ (ta, pu) ” denotes a pair in \mathcal{Y} with minimal similarity:

$$(17) s(ta, pu) \leq s(cv, pu) \text{ and } s(ta, pu) \leq s(c'v, bu) \text{ for all } c, c', v \text{ as in (16)}$$

Finally, choose any $v_2 \in V$ with $v_2 \neq a$ and $v_2 \neq u$; for ease of exposition, denote v_2 by i . Similarly, choose any $c_2 \in C$ with $c_2 \neq t$, $c_2 \neq p$ and $c_2 \neq b$; denote c_2 by k .

Now consider the following lexicon:

$$(18) \text{Lex}(L) \equiv \{ti, ka, pu, pu\}$$

Since t and a are present in the lexicon but b is not, Basic Combinatorial Induction (10) favors ta over bu :

$$(19) \text{Basic Combinatorial Induction} \Rightarrow \omega_L(ta) > \omega_L(bu)$$

Now in Lexical Similarity Theory (11d):

$$(20) \omega_L(ta) = f\{s(ta, ti), s(ta, ka), s(ta, pu), s(ta, pu)\}$$

where $f(\{\dots\})$ has been simplified to $f\{\dots\}$. But

$$(21) s(ta, ti) \leq s(bu, pu) \text{ and } s(ta, ka) \leq s(bu, pu) \text{ by (15), and } \\ s(ta, pu) \leq s(ti, bu) \text{ and } s(ta, pu) \leq s(ka, bu) \text{ by (17)}$$

Thus by the monotonicity of f (11e),

$$(22) f\{s(ta, ti), s(ta, ka), s(ta, pu), s(ta, pu)\} \leq f\{s(bu, pu), s(bu, pu), s(ti, bu), s(ka, bu)\}$$

But, by the symmetry of the similarity metric (11c), $s(ti, bu) = s(bu, ti)$, so the right side of (22) is exactly $\omega_L(bu)$; (22) and (20) thus give

$$(23) \omega_L(ta) \leq \omega_L(bu)$$

Since (23) contradicts (19), Lexical Similarity Theory is inconsistent with Basic Combinatorial Induction. \square

Acknowledgements

For helpful discussion of these and related issues I would like to thank Matt Goldrick, Adam Albright, Sharon Peperkamp, and especially Alan Prince, from whom I have had the great privilege and pleasure to learn so much on topics ranging from scientific explanation and empiricism to metrical theory, the RM logic, Hebbian learning, the eigenvalues of tridiagonal matrices, general relativity and beyond.

References

- Blevins, J. 1995. The syllable in phonological theory. In J. A. Goldsmith (ed.) *Handbook of Phonological Theory*, 206–244. Cambridge, MA: Blackwell.
- Coleman, J., and J. Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. In *Computational phonology: Third Meeting of the ACL Special Interest Group in Computational Phonology*, 49–56. Somerset, NJ: Association for Computational Linguistics Special Interest Group on Phonology (SIGPHON).
- Goldsmith, J. A. 1992. Local modeling in phonology. In S. Davis (ed.) *Connectionism: Theory and Practice*. Oxford, England: Oxford University Press.
- Goldsmith, J. A. 1993. Phonology as an intelligent system. In D. J. Napoli and J. Kegl (eds.), *Bridges between Psychology and Linguistics: A Swarthmore Festschrift for Lila Gleitman*, 247–267. Hillsdale, NY: Erlbaum.
- Goldsmith, J. A., and G. Larson. 1990. Local modeling and syllabification. *CLS* 26.
- Hayes, B., B. B. Tesar, and K. R. Zuraw. 2003. OTSoft 2.1. Software package. <http://www.linguistics.ucla.edu/people/hayes/otsoft/>.
- McCarthy, J. J., and A. Prince. 1993. Prosodic Morphology I: Constraint interaction and satisfaction. Technical report RuCCS-TR-3, Rutgers Center for Cognitive Science, Rutgers University, and University of Massachusetts at Amherst. [ROA-482, 2001, Rutgers Optimality Archive, <http://roa.rutgers.edu/>.]
- McCarthy, J. J., and A. Prince. 1995. Faithfulness and reduplicative identity. In J. Beckman, L. Walsh Dickey, and S. Urbanczyk (eds.), *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory*, 249–384. Amherst, MA: University of Massachusetts at Amherst, GLSA. [ROA-60, Rutgers Optimality Archive, <http://roa.rutgers.edu/>.]
- McClelland, J. L., and B. C. Vander Wyk. 2006. Graded constraints on English word forms. Ms.
- Pinker, S., and A. Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28, 73–193.
- Pinker, S., and A. Prince. 1996. The nature of human concepts: Evidence from an unusual source. *Communication and Cognition* 29, 307–362.
- Prince, A. 1993. In defense of the number *i*: Anatomy of a linear dynamic model of linguistic generalizations. Technical report RuCCS-TR-1, Rutgers Center for Cognitive Science, Rutgers University.
- Prince, A. 1996. Gradient ascent in a linear inhibitory network. RuCCS-TR-31, Rutgers Center for Cognitive Science, Rutgers University.
- Prince, A. 1997. Endogenous constraints on OT constraints. Talk presented at the Hopkins Optimality Theory Conference/Maryland Mayfest, Baltimore, MD.
- Prince, A. 2001. Invariance under re-ranking. Talk presented at the WCCFL 20, USC.
- Prince, A. 2002a. Anything goes. In T. Honma, M. Okazaki, T. Tabata, and S. Tanaka (eds.), *A New Century of Phonology and Phonological Theory*, 66–90. Tokyo: Kaitakusha. [ROA-536, Rutgers Optimality Archive, <http://roa.rutgers.edu/>.]
- Prince, A. 2002b. Arguing optimality. In A. Coetzee, A. Carpenter, and P. de Lacy (eds.), *Papers in Optimality Theory II*. Amherst, MA: UMass. GLSA. [ROA-536, Rutgers Optimality Archive, <http://roa.rutgers.edu/>.]

- Prince, A. 2002c. Entailed ranking arguments. [ROA-500, Rutgers Optimality Archive, <http://roa.rutgers.edu/>.]
- Prince, A. 2005. Harmony at base omega: Utility functions for OT. [ROA-798, Rutgers Optimality Archive, <http://roa.rutgers.edu/>.]
- Prince, A. 2006. The pursuit of theory. In P. de Lacy (ed.) *The Cambridge Handbook of Phonology*. Cambridge, England: Cambridge University Press.
- Prince, A., and S. Pinker. 1988a. Symbols aren't much good outside of a symbol-processing architecture. *Behavioral and Brain Sciences* 11, 46–47.
- Prince, A., and S. Pinker. 1988b. Wickelphone ambiguity. *Cognition* 30, 188–190.
- Prince, A., and P. Smolensky. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Technical report, Rutgers University and University of Colorado at Boulder, 1993. ROA 537, 2002. Revised version published by Blackwell, 2004.
- Prince, A., and P. Smolensky. 1997. Optimality: From neural networks to universal grammar. *Science* 275, 1604–1610.
- Prince, A., and P. Smolensky. 2003. Optimality Theory in phonology. In W. J. Frawley (ed.) *International Encyclopedia of Linguistics*. Oxford, England: Oxford University Press.
- Prince, A., and B. B. Tesar. 2004. Learning phonotactic distributions. In R. Kager, J. Pater, and W. Zonneveld (eds.), *Constraints in Phonological Acquisition*. Cambridge, England: Cambridge University Press. [ROA-353, Rutgers Optimality Archive, <http://roa.rutgers.edu/>.]
- Rumelhart, D. E., and J. L. McClelland. 1986. On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol. 2, *Psychological and Biological Models*, 216–271. Cambridge, MA: MIT Press.
- Samek-Lodovici, V., and A. Prince. 1999. Optima. [ROA-363, Rutgers Optimality Archive, <http://roa.rutgers.edu/>.]
- Samek-Lodovici, V., and A. Prince. 2002. Fundamental properties of Harmonic Bounding. [ROA-785, Rutgers Optimality Archive, <http://roa.rutgers.edu/>.]
- Smolensky, P. 2006. Optimality in phonology II: Harmonic completeness, local constraint conjunction, and feature domain markedness. In P. Smolensky and G. Legendre, *The Harmonic Mind: From Neural Computation to Optimality-theoretic Grammar*. Vol. 2, *Linguistic and Philosophical Implications*, 27–160. Cambridge, MA: MIT Press.
- Smolensky, P., and G. Legendre. 2006. *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Vol. 1. *Cognitive architecture*. Vol. 2. *Linguistic and philosophical implications*. Cambridge, MA: MIT Press.
- Tesar, B. B., J. Alderete, G. Horwood, N. Merchant, K. Nishitani, and A. Prince. 2003. Surgery in language learning. In *WCCFL* 22.
- Tesar, B. B., and A. Prince. 2003. Using phonotactics to learn phonological alternations. In *Chicago Linguistics Society: The panels* 39. [ROA-620, Rutgers Optimality Archive, <http://roa.rutgers.edu/>.]
- Vikner, S. 2001. V⁰-to-I⁰ movement and *do*-insertion in Optimality Theory. In G. Legendre, S. Vikner, and J. Grimshaw (eds.), *Optimality-Theoretic Syntax*, 427–464. Cambridge, MA: MIT Press.

Department of Cognitive Science
Johns Hopkins University
241B Krieger Hall / 3400 N. Charles Street
Baltimore, MD 21218-2685
smolensky@jhu.edu
<http://www.cogsci.jhu.edu/faculty/smolensky.html>

