Bruce Tesar & Paul Smolensky (2000) Learnability in Optimality Theory. Cambridge, Mass.: MIT Press. Pp. vii+140.

Reviewed by Paul Boersma, University of Amsterdam

January 17, 2004

This review is not only for people interested in learnability in OT, but for everybody who believes in OT's claim of *factorial typology*, i.e. the prediction that all language types generated by a permutation of the rankings of the constraints are attestable human languages. Learnability considerations are capable of shooting holes in that claim.

For the most part, the book is a slightly edited or re-edited collection of earlier work. Section 7.6 on Recursive Constraint Demotion and chapter 8 on production-directed parsing are from Tesar's 1995 dissertation, while chapters 1 through 3 and chapter 5 through section 7.5 constitute the authors' article in Linguistic Inquiry (1998), which again mainly recapitulated or copied Tesar's dissertation. Relatively new work is chapter 4, on "overcoming ambiguity in overt forms". This chapter is the latest and perhaps final version of a paper that appeared four times before (Tesar 1997, 1998, 1999, 2000), and unlike these earlier versions, it acknowledges the fact that the proposed learning algorithm *fails* for some possible and attested grammars.

The reader might think that when noting this record amount of recycling, I would advise against buying this book. The opposite is true. Anybody interested in learnability in OT (and anybody interested in possible holes in factorial typology) should buy this book and forget about reading the earlier work (i.e. Tesar & Smolensky 1993, 1996, 1998; Tesar 1995, 1997, 1998, 1999, 2000). This is possible because this book incorporates all relevant parts of that earlier work (obsolete versions of their learning algorithms, like Batch Constraint Demotion, are now ignored), and because this book does not redefine or reinterpret that earlier work at all. In other words, the book supersedes the earlier work by virtue of finishing it (chapter 4) or leaving it alone (the other chapters). Second, the book should be bought by anybody who thinks that

The oldest part of the learning algorithm presented in the book is Error-Driven Constraint Demotion (EDCD). Consider a child grammar as in (1). This is a *production tableau*, i.e., an underlying form is given, and a surface structure will have to be chosen.

(1) *Before EDCD (fully informed)*

$[\sigma\ \sigma\ \sigma\ \sigma]_{UF}$	TROCHAIC	IAMBIC	FEETRIGHT	FEETLEFT
[(σ́ σ) σ σ]		*	*!*	
√ [(σ σ́) σ σ]	*!		**	
[σ (σ́ σ) σ]		*	*!	*
[σ (σ σ́) σ]	*!		*	*
☞ [σ σ (σ́ σ)]		*		**
[σ σ (σ σ́)]	*!			**

The underlying form is [σ σ σ σ]$_{UF}$, i.e. a sequence of four syllables not marked for stress (we assume that this language has no lexical stress). Based on the ranking of the constraints, the output of grammar will be [σ σ (σ́ σ)], i.e. a right-aligned trochaic foot preceded by two extrametrical syllables. This is depicted by the pointing finger (☞). Now suppose that we tell the child that the actual form in the language she is trying to acquire is [(σ σ́) σ σ], i.e. a left-aligned iambic foot followed by two extrametrical syllables (this is possible if the adult grammar has the rankings IAMBIC >> TROCHAIC and FEETLEFT >> FEETRIGHT). This form happens to occur in the child's tableau, which we depict with a check mark (√). Now that the child's form and the adult form are different, EDCD will take action by looking up the highest ranked constraint that prefers the adult form to the learner's form (i.e. IAMBIC) and demoting all the higher ranked constraints that prefer the learner's form (in this case, only TROCHAIC) below this pivotal constraint. The constraint TROCHAIC thus ends up being ranked equally high as FEETRIGHT. The result is in tableau (2).

(2) *After first EDCD*

[σ σ σ σ]$_{UF}$	IAMBIC	TROCHAIC	FEETRIGHT	FEETLEFT
[(σ́ σ) σ σ]	*!		**	
√ [(σ σ́) σ σ]		*	*!*	
[σ (σ́ σ) σ]	*!		*	*
[σ (σ σ́) σ]		*	*!	*
[σ σ (σ́ σ)]	*!			**
☞ [σ σ (σ σ́)]		*		**

The adult form [(σ σ́) σ σ] has now become better in the child's grammar than the child's former form [σ σ (σ́ σ)], although a third form ([σ σ (σ σ́)]) has now become optimal in the child's grammar. The fact that the adult and child forms are still different means that we can apply EDCD again on the same underlying form. The pivotal constraint is FEETLEFT, and FEETRIGHT is the only constraint that has to be moved below it. This leads to Tableau 3. The child's form now equals the adult form, and EDCD will stop chewing on the underlying-surface pair [σ σ σ σ]$_{UF}$ – [(σ σ́) σ σ].

(3) *After second EDCD*

[σ σ σ σ]$_{UF}$	IAMBIC	TROCHAIC	FEETLEFT	FEETRIGHT
[(σ́ σ) σ σ]	*!			**
√☞ [(σ σ́) σ σ]		*		**
[σ (σ́ σ) σ]	*!		*	*
[σ (σ σ́) σ]		*	*!	*
[σ σ (σ́ σ)]	*!		**	
[σ σ (σ σ́)]		*	*!*	

The learner can now go on to learn from other underlying-surface pairs like [σ σ σ]$_{UF}$ – [(σ σ́) σ] and [σ σ σ σ σ]$_{UF}$ – [(σ σ́) σ σ σ]. But the grammar (3) already happens to work correctly for those pairs, since the ranking is that of an iambic left-aligning language. So this example was a quick success, in which only a single informative form was needed to establish the adult ranking. In general, however, EDCD will require several different pairs of underlying forms and surface forms before homing in on the target language. In the book, T&S show that if the learner is given enough randomly selected pairs of underlying form and surface structure, EDCD will eventually change the grammar in such a way that it assigns correct surface structures to all underlying forms. In other words, EDCD is guaranteed to succeed when full structural descriptions of surface forms are given.

But full structural descriptions of surface forms are not generally provided to the learner. In reality, the child only hears the overt form [σ σ́ σ σ]$_{OF}$, not the phonological structure [(σ σ́) σ σ], i.e., she may be able to hear which syllable is stressed but the foot structure is hidden. T&S now propose that the learner uses her OT grammar to infer a surface structure from the overt form. Thus, if the constraint ranking is as in (1), i.e. trochaic right-aligning, the learner will interpret the overt form [σ σ́ σ σ]$_{OF}$ as the phonological structure [σ (σ́ σ) σ]. The interpretation tableau (3) shows how this works.

(4) *Grammar-guided interpretation by the learner*

[σ σ́ σ σ]$_{OF}$	TROCHAIC	IAMBIC	FEETRIGHT	FEETLEFT
[(σ σ́) σ σ]	*!		**	
☞ [σ (σ́ σ) σ]		*	*	*

T&S call this *Robust Interpretive Parsing* (RIP): the learner will assign to the overt form a structure that minimally violates her constraint ranking, even if this structure is ungrammatical in her own production (as it is here, because she would pronounce an underlying [σ σ σ σ]$_{UF}$ as [σ σ (σ́ σ)]). This makes a large difference for EDCD, which is now given not the correct adult form, but a possibly incorrect form that results from the child's interpretation. Instead of (1), we now have Tableau (5).

(5) *Before EDCD (partly informed)*

[σ σ σ σ]$_{UF}$	TROCHAIC	IAMBIC	FEETRIGHT	FEETLEFT
[(σ́ σ) σ σ]		*	*!*	
[(σ σ́) σ σ]	*!		**	
√ [σ (σ́ σ) σ]		*	*!	*
[σ (σ σ́) σ]	*!		*	*
☞ [σ σ (σ́ σ)]		*		**
[σ σ (σ σ́)]	*!			**

EDCD will demote FEETRIGHT below FEETLEFT. This leads to Tableau (6), where [(σ́ σ) σ σ] has become the child's form.

(6) *After EDCD (partly informed)*

[σ σ σ σ]$_{UF}$	TROCHAIC	IAMBIC	FEETLEFT	FEETRIGHT
☞ [(σ́ σ) σ σ]		*		**
[(σ σ́) σ σ]	*!			**
[σ (σ́ σ) σ]		*	*!	*
[σ (σ σ́) σ]	*!		*	*
[σ σ (σ́ σ)]		*	*!*	
[σ σ (σ σ́)]	*!		**	

The learner can try to proceed by re-chewing the same overt form [σ σ́ σ σ]$_{OF}$. The interpretation is again [σ (σ́ σ) σ], because in (6) that form is better than its only competitor [(σ σ́) σ σ]. The learner will therefore demote FEETLEFT below FEETRIGHT (imagine a check mark for the third candidate), and return to the situation in (5). This is an example of the general way in which RIP/EDCD can get stuck: because of a non-adultlike interpretation of overt forms, the learner will rerank the wrong constraints and end up visiting an eternal cycle of inadequate grammars, i.e. grammars that produce a non-adultlike overt form for at least one underlying form. In the example at hand, no sequence of adult overt forms [σ σ́ σ]$_{OF}$, [σ σ́ σ σ]$_{OF}$, and [σ σ́ σ σ σ]$_{OF}$ will lead to a correct reranking of the two foot form constraints TROCHAIC and IAMBIC; only if the language contains the disyllabic form [σ σ́]$_{OF}$ will the learner be able to rank these constraints correctly and ultimately come up with an adequate grammar.

This is the most interesting point that T&S make. The algorithm (RIP/CD) can fail to converge on a correct ranking for the target language. Of course, the success of the algorithm will depend on the initial ranking of the constraints. T&S describe a computer simulation of a metrical stress example with 12 constraints, starting with all constraints ranked at the same height. They fed RIP/CD with overt forms from any of 124 artificial languages. Only 75 of those languages were learned correctly by RIP/CD. When starting with high-ranked constraints for foot form (trochaicity and iambicity), the number of successes rose to 94. When the weight-to-stress principle was initially raised even above the foot form constraints, the number rose to 120. Your reviewer is somebody who wants to check such claims, so I recreated T&S's metrics grammar in the Praat program (www.praat.org), giving 62 tableaus with a total of 15344 candidates and 370404 violation marks. I then taught 10 groups of 124 virtual learners the 124 languages. Although the overt forms were presented to each learner in the same order (beginning with the shortest forms), the 10 learners of a particular language sometimes performed in slightly different ways, because in interpretation several candidates may tie for optimality, and the learner has to randomly choose from them. With an equal initial ranking, the average number of successful learners was 72.1, with an initial high ranking of the foot form constraints, it was 92.1, and with an initial high ranking of weight-to-stress, it was 114.5. These averages are slightly lower than the ones reported by T&S, probably because of a different handling of ties in interpretation (Tesar p.c. says that when two candidates tied for optimality in their simulations, T&S deterministically but unrealistically chose the one that happened to occur earlier in the tableau).

The high performance of RIP/CD reported by T&S depends on several assumptions. To see this, I considered 36 supergroups of 10 groups of 124 learners. The supergroups varied with respect to the order of presentation of the overt forms: for one third of the learners the forms were presented in T&S's cyclically applied fixed short-to-long-order, one third heard each cycle of 62 overt forms in randomly permuted order, and one third heard the data randomly drawn from the 62 possible forms, which arguably resembles best the actual acquisition process. The supergroups also varied with respect to the handling of tied constraints: one half of the learners allowed crucial ties, i.e. constraints whose violation marks count together if they are ranked at the same height, as in T&S's simulations; since crucial ties may be unrealistic (for how can one weigh a single violation of the binary constraint NONFINALITY against multiple violations of the gradient constraint FEETLEFT?), the other half of the learners had the variationist interpretation of tied constraints (Anttila 1997), in which constraints are randomly ordered at each evaluation if they have the same ranking (this can be simulated in Praat by using a tiny bit of evaluation noise). Finally, the supergroups varied with respect to how often every single datum was processed: one half of the learners was allowed to chew five times on each language datum, and to backtrack if this form did not become grammatical (T&S: 69); the other half interpreted and reproduced each datum only once, with no backtracking. It turned out that on average: (1) a group of 124 learners who hear the fixed order acquires 6 more languages than a group of 124 learners who hear randomized or random orders; (2) learners with crucial ties acquire 12 more languages than those with variationist ties; (3) the number of chews did not have any effect on acquisition performance; and (4) learners with the foot-form-high initial state acquired 17 languages more than learners with an equally-ranked initial state, but 18 languages fewer than learners with an initial state in which weight-to-stress was ranked even higher.

It is not generally bad for a learning algorithm to fail on certain input data. If an OT learning algorithm can predict that certain constraint rankings are unlearnable, and exactly these rankings turn out not to occur in the languages of the world, such an unlearnability result constitutes positive support for that algorithm (Clark & Roberts 1993). Any of the following things may influence learnability: the constraint set (according to Apoussidou & Boersma 2003, there is no ranking of T&S's 12 constraints that can describe Latin stress if there is main stress only), the initial hierarchy (as T&S show), the order of presentation of the data (perhaps the kids pay attention to shorter forms first), or details of the learning algorithm. As an example of the latter, pride forced me to have a look at what the performance would be if EDCD is replaced with GLA (Boersma & Hayes 2001), an algorithm that indiscriminately promotes the ranking of all the constraints that prefer the adult form over the child's form (in (1) these are IAMBIC and FEETLEFT) and demotes the ranking of all the constraints that prefer the child's form over the adult form (in (1) these are TROCHAIC and FEETRIGHT). Compared to RIP/CD with randomly drawn language data and variationist ties, a group of 124 RIP/GLA learners acquired 11 languages more (averaged over the three different initial states), which is comparable to the performance of RIP/CD with crucial ties. It is clear that much research is needed to find out whether there is any combination of algorithms, constraint sets, and initial states that accurately predicts the learnability of attested languages and at the same time is capable of showing that many attested gaps in the factorial typology are not accidental but can be explained by the formal unlearnability of such languages. An example of such research is Jäger (to appear), who goes even further by predicting not only some learnable and unlearnable languages but also some learnable but diachronically unstable languages.

While the typological gaps predicted by the metrical examples of chapter 4 are the result of complicated constraint interactions and therefore hard to explain, chapter 5 contains an example from which we can easily produce a prediction of a straightforward typological gap. In §5.2 T&S discuss the learning of underlying forms by means of lexicon optimization. Their example is the German pair of overt forms $[\text{tak}]_{OF}$ 'day' - $[\text{tagə}]_{OF}$ 'days'. Since German has voice neutralization word-finally and a voicing contrast intervocalically, the underlying forms must be $[\text{tag+}\varnothing]_{UF}$ and $[\text{tag+ə}]_{UF}$, and a possible OT explanation for the surface forms is Lombardi's (1999) ranking ONSETFAITH >> *[+voi] >> FAITH. T&S show that the underlying forms are learnable. In a lexicon optimization tableau, several candidate underlying-surface pairs are compared, under the condition that the surface forms share a given overt form. The paradigm $[\text{tag+}\varnothing]_{UF}$ - $[\text{tag+ə}]_{UF}$ with the surface forms $[\text{tag}_{\text{<voi>}}]_{SF}$ - $[\text{tagə}]_{SF}$ violates FAITH only (final devoicing), whereas the paradigm $[\text{tak+}\varnothing]_{UF}$ - $[\text{tak+ə}]_{UF}$ with $[\text{tak}]_{SF}$ - $[\text{tak}_{\text{[voi]}}\text{ə}]_{SF}$, violates both ONSETFAITH and FAITH (intervocalic voicing). Note that the constraint *[+voi] evaluates surface forms only and cannot decide between the two paradigm candidates. But now consider a language that I will call 'anti-German', a hypothetical language with final voicing contrasts but with intervocalic voicing. Such a language is predicted by a ranking like *V[–voi]V >> ONSETFAITH >> FAITH >> *[+voi]. The overt pair $[\text{tak}]_{OF}$ - $[\text{tagə}]_{OF}$ can only derive from the underlying forms $[\text{tak+}\varnothing]_{UF}$ - $[\text{tak+ə}]_{UF}$ with surface structures $[\text{tak}]_{SF}$ - $[\text{tak}_{\text{[voi]}}\text{ə}]_{SF}$. However, lexicon optimization can again only propose the paradigm $[\text{tag+}\varnothing]_{UF}$ and $[\text{tag+ə}]_{UF}$ with $[\text{tag}_{\text{<voi>}}]_{SF}$ - $[\text{tagə}]_{SF}$, independently of the ranking of ONSETFAITH and FAITH, because the two structural constraints again make no difference between the candidates. Thus, T&S' version of lexicon optimization predicts that anti-German is unlearnable and that it constitutes a gap in factorial typology. Now suppose that anti-German does not exist. It used to be the case that such a situation was regarded as evidence against this set of four constraints. But now we know that the non-existence of anti-German is actually predicted by learnability issues although the constraint set could be fine. Of course, if anti-German turns out to exist after all, either the constraint set or the learning algorithm must be wrong. Unfortunately, T&S appear to be concerned only with maximizing the learning scores of their algorithms. Never do they themselves draw the conclusion that a failure to learn can point to a genuine gap in factorial typology. Still, such should be standpoint of anybody who has up to this day, when confronted with the lack of attestation of a language type predicted by factorial typology, drawn the conclusion that there must be something wrong with the constraint set. Rather than insisting on a constraint set that produces the precise attested typology under ranking permutation, more OT researchers should start to take into account the possibility that some typological gaps could be caused by a lack of learnability.

And now for some minor critical remarks. For child language researchers who want to dive into formal learnability, the book presents some confusing terminology. T&S use the term "input" in the sense of 'underlying form', and the term "output" in the sense of 'fully structured surface form'. This is traditional OT usage, and appropriate when we talk about production only, but the terms are unfortunate when we talk about interpretation, in which case the overt form should be regarded as the "input". In chapter 4, where T&S take overt forms into account, the terms "input" and "output" are correctly replaced with Underlying Form and Full Structural Description, but the problematic terms still occur in most of the book, simply because of the slightness of the editing. In the child language literature, moreover, the term "input" is used in the sense of 'primary language data', which is the same as what T&S call the 'overt form'. It may be true that language acquisition researchers often regard the child's underlying form as identical to the overt adult form, but that does not make these forms *the same.* Another problematic pair of terms is "loser" versus "winner". T&S call the candidate that is optimal in the learner's grammar the "loser", and the form that the learner considers to be the correct adult form the "winner", as if the learner is taking an adult standpoint when judging the appropriateness of her grammar; in the more common child-centred approach, the two terms clearly need to be reversed. A third problematic term is "interpretive parsing", i.e. the mapping from overt to surface form. This process is known among psycholinguists and phoneticians as "perception". The renaming may have been due partly to a shyness against terms that sound extragrammatical, partly to the tacit understanding that the mapping from surface form to underlying form (what others call "recognition") is part of the interpretive parsing stage.

With this discussion of the terms for the stages of comprehension, we arrive at another problematic theoretical issue, namely the authors' reliance on the *containment* model of Optimality Theory. This model assumes that both the underlying form and the overt form are *contained* in the full structural description or can be trivially derived from this surface structure. For instance, the containment view of the finally-devoiced surface form of the German underlying form $[\text{tag}]_{\text{UF}}$ is $[\text{tag}_{<\text{voi}>}]$. McCarthy & Prince (1995) replaced this model by the correspondence view of faithfulness, in which this surface form is just [tak]. This obviously renders the surface-to-underlying mapping non-trivial. The same fate must hit the surface-to-overt mapping (i.e. phonetic implementation), which is language-dependent (e.g. stress is implemented by different cues in different languages), hence non-trivial as well. The extension of RIP/CD to these more comprehensive theories of phonological representation seems to have to involve a major future research effort.

The slightness of the authors' editing furthermore led to conflicting remarks about the implications of their learning algorithms for the *subset problem*, i.e. the problem that a learning algorithm may lead to a *superset language*, a language that consists of all possible adult forms *and some more*. On page 76, they claim that this can be solved by assuming an initial ranking of structural over faithfulness constraints. On page 100, they grant that sometimes EDCD "may converge on a superset language". Crucially, however, page 110 explicitly states that EDCD automatically generates informative "losers", i.e., that the negative evidence needed to get out of a superset language is provided by the learner herself. To solve the riddle posed by these three disparate remarks, the reader would have to work out by herself that the first two claims refer to the *covert* subset problem, the situation in which the grammar would allow unattested forms (e.g. [CVC] structures) for possibly non-occurring underlying forms (e.g. /CVC/ in a no-coda language), as required by Richness of the Base, and that the third claim refers to the overt subset problem, the situation in which the learner actually produces forms that are ungrammatical in the adult language. The overt subset problem referred to here only appears if some underlying forms have multiple optimal outputs in the superset language. The discussion about this in Tesar (1995: §4.6.2) is the only relevant earlier work that did not make it into this book. I can think of two reasons for that: first, a discussion of optionality would require reference to later work by others (e.g. Boersma & Hayes 2001), violating the authors' slight-editing principle; secondly, Tesar's (1995) solution was that optionality did not exist, which is a standpoint not necesarily shared with the other author.

T&S claim that OT learning algorithms are specific to the language faculty, unlike P&P (Principles & Parameters) learning algorithms. This claim is argued for by the assertion that P&P learning algorithms could equally well be applied to problems outside linguistics, such as 'training a neural network (with binary weights) to classify radar images of submarines' (p. 2), and by the assertion that OT learning algorithms have no application outside linguistics. But it seems to me that P&P and OT do not differ that much in their applicability outside linguistics. In fact, OT can be applied as a general decision scheme. For instance, OT seems to be the natural framework for describing the ranking of what are called 'rules' in everyday life. The ordering of traffic rules is a good example [until recently, the Dutch ranking for the right of way would have been: police person's directions >> ambulance >> traffic lights >> pedestrian crossing >> { straight on or tram } >> priority sign (non-pedestrian) >> tram >> car >> coming from the right >> bicycle]. EDCD would work perfectly when fed with traffic situations while being told who has priority. Finally, I am sure that if OT can be used for classifying phonological feature values (e.g. Escudero & Boersma 2003), it can just as well be used for classifying radar images.

In all, this book has been and will be the starting point for all subsequent work in the modelling of actual acquisition data (e.g. Curtin & Zuraw 2001), the modelling of learnability in more comprehensive views of the grammar (e.g. Escudero & Boersma 2003), the modelling of covert subset phenomena (e.g. Hayes to appear, and Prince & Tesar 1999), the modelling of optionality (e.g. Boersma & Hayes 2001), and the modelling of language change (e.g. Jäger to appear). For the general OT phonologist, the failures of the learning algorithms noted in this book and the failures predicted in this review should be a warning that the connection between constraint ranking and typology cannot be as intimate as was claimed in the original papers that defined Optimality Theory. Children have to learn their languages from incomplete representations of adult linguistic structures, and it is likely that this incompleteness poses large restrictions on what types of languages are possible and what types are not.

Anttila, Arto (1997). *Variation in Finnish phonology and morphology*. PhD thesis, Stanford University.

Apoussidou, Diana & Paul Boersma (2003). The learnability of Latin stress. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* **25**: xx–xx.

Boersma, Paul & Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* **32**: 45–86.

Clark, Robin & Ian Roberts (1993). A computational model of language learnability and language change. *Linguistic Inquiry* **24**: 299–345.

Curtin, Suzanne, & Kie Zuraw (2002). Explaining constraint demotion in a developing system. In Barbora Skarabela, Sarah Fish, and Anna H.-J. Do (eds.) *Proceedings of the 26th Annual Boston University Conference on Language Development*. Somerville, Mass.: Cascadilla Press. 118–129.

Escudero, Paola, & Paul Boersma (2003). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In Sudha Arunachalam, Elsi Kaiser & Alexander Williams (eds.): *Proceedings of the 25th Annual Penn Linguistics Colloquium. Penn Working Papers in Linguistics* **8.1**: 71–85. [ROA **439**, 2001]

Hayes, Bruce (to appear). *Phonological acquisition in OT: the early stages*. In René Kager, Joe Pater & Wim Zonneveld (eds.): *Fixing priorities: constraints in phonological acquisition*. Cambridge University Press. [ROA **327**, 1999]

Jäger, Gerhard (to appear). Learning constraint sub-hierarchies: The Bidirectional Gradual Learning Algorithm. In Henk Zeevat & Reinhard Blutner (eds.) *Optimality Theory and Pragmatics*. Palgrave Macmillan. [ROA **544**, 2002]

Lombardi, Linda (1999). Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* **17**: 267–302.

Prince, Alan & Bruce Tesar (1999). *Learning phonotactic distributions*. Technical Report TR-**54**, Rutgers University Center for Cognitive Science. [Rutgers Optimality Archive **353**]

Tesar, Bruce (1995). *Computational Optimality Theory*. PhD thesis, University of Colorado. [ROA **90**]

Tesar, Bruce (1997). An iterative strategy for learning metrical stress in Optimality Theory. In Elizabeth Hughes, Mary Hughes, & Annabel Greenhill (eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development*, 615–626. Somerville, Mass.: Cascadilla. [ROA **177**]

Tesar, Bruce (1998). An iterative strategy for language learning. *Lingua* **104**: 131–145.

Tesar, Bruce (1999). Robust interpretive parsing in metrical stress theory. In Kimary Shahin, Susan Blake & Eun-Sook Kim (eds.), *Proceedings of the 17th West Coast Conference on Formal Linguistics*, 625–639. Stanford, Calif.: CSLI. [ROA **262**]

Tesar, Bruce (2000). On the roles of optimality and strict domination in language learning. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.), *Optimality Theory: Phonology, syntax, and acquisition*, 592–620. New York: Oxford University Press.

Tesar, Bruce & Paul Smolensky (1993). *The learnability of Optimality Theory: an algorithm and some basic complexity results*. Ms. Department of Computer Science & Institute of Cognitive Science, University of Colorado at Boulder. [ROA **2**]

Tesar, Bruce & Paul Smolensky (1996). *Learnability in Optimality Theory (long version)*. Technical Report **96-3**, Department of Cognitive Science, Johns Hopkins University, Baltimore. [ROA **156**]

Tesar, Bruce & Paul Smolensky (1998). Learnability in Optimality Theory. *Linguistic Inquiry* **29**: 229–268.