Phonology **20** (2003) 436–446. © 2003 Cambridge University Press
DOI: 10.1017/S0952675704230111 Printed in the United Kingdom

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

Paul Boersma
University of Amsterdam

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. As the study of formal models of language acquisition, learnability theory is capable of making precise predictions not only about how children acquire the same language as their parents, but also about acquisition-induced language change (e.g. Clark & Roberts 1993, Lightfoot 1999, Jäger 2003), with typological consequences. Although the authors do not discuss such issues, I will address in this review a couple of typological points that can be extracted directly or indirectly from the book.

For the most part, the book is a slightly edited or re-edited collection of earlier work. §7.6 on Recursive Constraint Demotion and Chapter 8 on production-directed parsing are from Tesar's 1995 dissertation, while Chapters 1–3 and 5–7 (up to §7.5) constitute the authors' article in *Linguistic Inquiry* (1998), which again mainly recapitulated or copied Tesar (1995). Chapter 4, 'Overcoming ambiguity in overt forms', is relatively new work. This chapter is the latest and perhaps final version of a paper that appeared four times earlier in various forms (Tesar 1997, 1998, 1999, 2000); unlike these earlier versions, it acknowledges the fact that the proposed learning algorithm *fails* for some possible and perhaps attested grammars.

The reader might think that in noting this record amount of recycling, I am advising 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 it 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 ignored), and because it 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 as it was (the other chapters).

The oldest part of the acquisition model presented in the book is Error-Driven Constraint Demotion (EDCD). This is an algorithm capable of turning a current child grammar (i.e. an intermediate stage at some point during the acquisition process) into a new child grammar (the next stage) on the basis of a single piece of incoming language data. 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 (since we will have to distinguish three kinds of representations, I will denote underlying forms with pipes, surface structures with slashes and overt forms with square brackets).

(1) *Before EDCD* (*fully informed*)

|σ σ σ σ|	Trochaic	Iambic	FeetR	FeetL
a. /(σ́ σ) σ σ/		*	*!*	
✓ b. /(σ σ́) σ σ/	*!		**	
c. /σ (σ́ σ) σ/		*	*!	*
d. /σ (σ σ́) σ/	*!		*	*
☞ e. /σ σ (σ́ σ)/		*		**
f. /σ σ (σ σ́)/	*!			**

The underlying form is |σ σ σ σ|, 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 the grammar will be /σ σ (σ́ σ)/, i.e. a right-aligned trochaic foot preceded by two extrametrical syllables. This is represented by the familiar pointing finger. Now suppose (unrealistically, as we will see) that we tell the child explicitly 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 FeetL ≫ FeetR). This form happens to occur in the child's tableau, which we represent 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 even higher-ranked constraints that prefer the learner's form (in this case, only Trochaic) below this pivotal constraint. The constraint Trochaic thus ends up with the same ranking as FeetR. The resulting new grammar is in tableau (2).

(2) *After first EDCD*

|σ σ σ σ|	Iambic	Trochaic	FeetR	FeetL
a. /(σ́ σ) σ σ/	*!		**	
✓ b. /(σ σ́) σ σ/		*	*!*	
c. /σ (σ́ σ) σ/	*!		*	*
d. /σ (σ σ́) σ/		*	*!	*
e. /σ σ (σ́ σ)/	*!			**
☞ f. /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)/) is now 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 FeetL, and FeetR is the only constraint that has to be moved below it. This leads to tableau (3). The child's form is now identical to the adult form, and EDCD will stop chewing on the underlying–surface pair |σ σ σ σ| – /(σ σ́) σ σ/.

(3) *After second EDCD*

$\|\sigma\,\sigma\,\sigma\,\sigma\|$	IAMBIC	TROCHAIC	FEETL	FEETR
a. $/(\acute{\sigma}\,\sigma)\,\sigma\,\sigma/$	*!			**
✓☞ b. $/(\sigma\,\acute{\sigma})\,\sigma\,\sigma/$		*		**
c. $/\sigma\,(\acute{\sigma}\,\sigma)\,\sigma/$	*!		*	*
d. $/\sigma\,(\sigma\,\acute{\sigma})\,\sigma/$		*	*!	*
e. $/\sigma\,\sigma\,(\acute{\sigma}\,\sigma)/$	*!		**	
f. $/\sigma\,\sigma\,(\sigma\,\acute{\sigma})/$		*	*!*	

The learner can now go on to learn from other underlying–surface pairs like $|\sigma\,\sigma\,\sigma| - /(\sigma\,\acute{\sigma})\,\sigma/$ and $|\sigma\,\sigma\,\sigma\,\sigma\,\sigma| - /(\sigma\,\acute{\sigma})\,\sigma\,\sigma\,\sigma/$. But the grammar in (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, Tesar & Smolensky 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 $[\sigma\,\acute{\sigma}\,\sigma\,\sigma]$, not the phonological structure $/(\sigma\,\acute{\sigma})\,\sigma\,\sigma/$, i.e. she may be able to hear which syllable is stressed, but the foot structure is hidden. Tesar & Smolensky now propose that the learner uses her OT grammar to infer a surface structure from the overt form. Thus, if her current constraint ranking is as in (1), i.e. trochaic right-aligning, the learner will interpret the overt form $[\sigma\,\acute{\sigma}\,\sigma\,\sigma]$ as the phonological structure $/\sigma\,(\acute{\sigma}\,\sigma)\,\sigma/$. The INTERPRETATION TABLEAU (4) shows how this works (both candidates have stress on the second syllable, so that they are compatible with the given overt form).

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

$[\sigma\,\acute{\sigma}\,\sigma\,\sigma]$	TROCHAIC	IAMBIC	FEETR	FEETL
a. $/(\sigma\,\acute{\sigma})\,\sigma\,\sigma/$	*!		**	
☞ b. $/\sigma\,(\acute{\sigma}\,\sigma)\,\sigma/$		*	*	*

Tesar & Smolensky 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 produce an underlying $|\sigma\,\sigma\,\sigma\,\sigma|$ as $/\sigma\,\sigma\,(\acute{\sigma}\,\sigma)/$). 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 (5).

(5) *Before EDCD* (*partly informed*)

|σ σ σ σ|	Trochaic	Iambic	FeetR	FeetL
a. /(σ́ σ) σ σ/		*	*!*	
b. /(σ σ́) σ σ/	*!		**	
✓ c. /σ (σ́ σ) σ/		*	*!	*
d. /σ (σ σ́) σ/	*!		*	*
☞ e. /σ σ (σ́ σ)/		*		**
f. /σ σ (σ σ́)/	*!			**

EDCD will demote FeetR below FeetL. 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|	Trochaic	Iambic	FeetL	FeetR
☞ a. /(σ́ σ) σ σ/		*		**
b. /(σ σ́) σ σ/	*!			**
c. /σ (σ́ σ) σ/		*	*!	*
d. /σ (σ σ́) σ/	*!		*	*
e. /σ σ (σ́ σ)/		*	*!*	
f. /σ σ (σ σ́)/	*!		**	

The learner can try to proceed by processing the same overt form [σ σ́ σ σ] again. The interpretation is again /σ (σ́ σ) σ/, as in (4), since Trochaic is still top-ranked. The learner will therefore demote FeetL below FeetR (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/CD (i.e. RIP with EDCD) can get stuck: because of a non-adult-like 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-adult-like overt form for at least one underlying form. In the example at hand, no sequence of adult overt forms [σ σ́ σ], [σ σ́ σ σ], [σ σ́ σ σ σ] 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] 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 Tesar & Smolensky 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. Tesar & Smolensky describe a computer simulation of a metrical stress example with twelve constraints, starting with all constraints having the same ranking. They fed RIP/CD with 62 overt forms (from two to seven syllables, with varying syllable weights and varying main and secondary stresses) from each 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 above even the foot-form constraints, the number rose to 120. Your reviewer is somebody who wants to check such claims, so I recreated Tesar & Smolensky's metrics grammar in the Praat program (www.praat.org), giving 62 tableaux, with a total of 15,344 candidates and 370,404 violation marks. I then taught ten groups of 124 virtual learners the same 124 languages as Tesar & Smolensky did (Bruce Tesar kindly supplied the 124 sets of 62 overt forms). Although the overt forms were presented to every learner in the same order (beginning with the shortest forms), the ten 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 choose randomly 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 Tesar & Smolensky, probably because of a different handling of ties in interpretation (Tesar (personal communication) says that when two candidates tied for optimality in their simulations, Tesar & Smolensky deterministically but unrealistically chose the one that happened to occur earlier in the tableau).

The high performance of RIP/CD reported by Tesar & Smolensky depends on several assumptions. To see this, I considered 36 supergroups of ten 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 Tesar & Smolensky'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 Tesar & Smolensky's simulations; since crucial ties may be unrealistic (for how can one weigh a single violation of the binary constraint NON-FINALITY against multiple violations of the gradient constraint FEETL?), 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). The supergroups furthermore varied with respect to their initial rankings, as before: one third of the learners started with equal rankings, one third with high foot-form constraints and one third with even higher weight-to-stress. 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 (Tesar & Smolensky, p. 69); the other half interpreted and reproduced each datum only once, with no backtracking. It turned out that on average: (i) a group of 124 learners who hear the fixed order acquires six more languages than a group of 124 learners who hear randomised or random orders; (ii) learners with crucial ties acquire twelve more languages than those with variationist ties; (iii) the number of chews on each language datum did not have any effect on acquisition performance; and (iv) learners with the foot-form-high initial state acquired seventeen languages more than learners with an equally ranked initial state, but eighteen 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. In fact, Eisner (2000) shows that such failures must be expected from *any* OT learning algorithm when the language data are only partially known (as is the case with overt forms): Eisner showed that for any learning algorithm and every size of the constraint set one can construct a combination of constraints, candidates and a surface-to-overt mapping for which the learning algorithm, in order to be 100% successful, would not be qualitatively faster than enumerating all possible rankings, which is impractical for real language learners. If, then, 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. Any of the following things may influence learnability: the constraint set (according to Apoussidou & Boersma 2003, there is no ranking of Tesar & Smolensky's twelve constraints that can describe Latin stress if there is main stress only), the initial hierarchy (as Tesar & Smolensky show), the order of presentation of the data (perhaps the children 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 FEETL) and demotes the ranking of all the constraints that prefer the child's form over the adult form (in (1) these are TROCHAIC and FEETR). Compared to RIP/CD with randomly drawn language data and variationist ties, a group of 124 RIP/GLA learners acquired eleven 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 (2003), 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 Tesar & Smolensky discuss the learning of underlying forms by means of lexicon optimisation. Their example is the German overt paradigm [tak ~ tagə] 'day ~ days'. Since German has voice neutralisation word-finally and a voicing contrast intervocalically, the underlying paradigm must be |tag + ∅ ~ tag + ə|, and a possible OT explanation for the surface forms is Lombardi's (1999) ranking ONSETFAITH ≫ *[+voi] ≫ FAITH, as shown in (7) (subscript <voi> in the tableaux below indicates that voice has been deleted; subscript [voi] that it has been inserted).

(7) *Final devoicing in German: production*

|tag+∅ ~ tag+ə|	OnsetFaith	*[+voi]	Faith
a. /tag ~ tagə/		**!	
☞ b. /tag$_{<voi>}$ ~ tagə/		*	*
c. /tag$_{<voi>}$ ~ tag$_{<voi>}$ə/	*!		**

Tesar & Smolensky show that the underlying paradigm is learnable. In the LEXICON-OPTIMISATION TABLEAU (8), several candidate underlying–surface pairs are compared, under the condition that the surface paradigms share the given overt paradigm [tak ~ tagə].

(8) *Final devoicing in German: lexicon optimisation*

[tak ~ tagə]	OnsetFaith	*[+voi]	Faith
☞ a. |tag+∅ ~ tag+ə| – /tag$_{<voi>}$ ~ tagə/		*	*
b. |tak+∅ ~ tak+ə| – /tak ~ tak$_{[voi]}$ə/	*!	*	*

The first pair of underlying and surface paradigms violates Faith (final devoicing), whereas the second pair violates both OnsetFaith and Faith (intervocalic voicing). Note that the constraint *[+voi] evaluates surface forms only, and thus has no preference for either of the two paradigm candidates. The first candidate, which is appropriate for German, wins, independently of the ranking of the constraints, so that we can conclude that this German alternation is learnable. 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 the one in tableau (9).

(9) *Intervocalic voicing but no final devoicing in anti-German: production*

	|tak+∅ ~ tak+ə|	*V[−voi]V	OnsetFaith	Faith	*[+voi]
a.	i. /tak ~ takə/	*!			
	☞ ii. /tak ~ tak$_{[voi]}$ə/		*	*	*
	iii. /tak$_{[voi]}$ ~ tak$_{[voi]}$ə/		*	**!	**
b.	|tag+∅ ~ tag+ə|				
	☞ i. /tag ~ tagə/				**
	ii. /tag$_{<voi>}$ ~ tagə/			*!	*
	iii. /tag$_{<voi>}$ ~ tag$_{<voi>}$ə/	*!	*	**	

We see that in this language the overt paradigm [tak ~ tagə] can only derive from the underlying paradigm |tak + ∅ ~ tak + ə|. However, as tableau (10) shows, lexicon optimisation can again only propose the underlying paradigm |tag + ∅ ~ tag + ə|, independently of the ranking of the constraints, because the two structural constraints again make no difference between the candidates.

(10) *Anti-German: lexicon optimisation fails*

[tak ~ tagə]	*V[−voi]V	OnsetFaith	Faith	*[+voi]
☞ a. |tag+∅ ~ tag+ə| – /$\text{tag}_{\text{<voi>}}$ ~ tagə/			*	*
b. |tak+∅ ~ tak+ə| – /tak ~ $\text{tak}_{\text{[voi]}}$ə/		*!	*	*

Thus, Tesar & Smolensky's version of lexicon optimisation must predict that all learners come up with a German analysis of anti-German, and hence that anti-German cannot exist for more than a single generation, so that it must constitute a gap in factorial typology. Now suppose that anti-German, a language with intervocalic voicing but without final devoicing, 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 be a stable existing language after all, either the constraint set or the acquisition model must be wrong. Unfortunately, Tesar & Smolensky appear to be concerned only with maximising 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. But this should be the standpoint of anybody who, when confronted with the lack of attestation of a language type predicted by factorial typology, has 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. Tesar & Smolensky 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 are talking only about production, but the terms are unfortunate when we are talking about interpretation, in which case the overt form should be regarded as the 'input'. In Chapter 4, where Tesar & Smolensky 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 Tesar & Smolensky 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' *vs.* 'winner'. Tesar & Smolensky 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 about using terms that sound extra-grammatical, partly to the tacit understanding that the mapping from surface form to underlying form (what others call 'recognition') could be part of the interpretive parsing stage, and partly by the need to find a term that would cover the analogous process in syntax.

This discussion of the terms for the stages of comprehension brings us to another problematic theoretical issue, 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 |tag| 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 befall the surface-to-overt mapping (i.e. phonetic implementation), which is language-dependent (e.g. stress is implemented by different cues in different languages), and hence non-trivial as well. The extension of RIP/CD to these more comprehensive theories of phonological representation will involve a major research effort in the future.

The slightness of the authors' editing leads 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 this: first, a discussion of optionality would require reference to later work by others (e.g. Boersma 1997), violating the authors' slight-editing principle; secondly, Tesar's (1995) solution was that optionality did not exist, which is a standpoint not necessarily shared with the other author.

Tesar & Smolensky 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 assertions 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. 4), and 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 right of way was: being directed by the police $\gg$ being (in) an ambulance $\gg$ having a green traffic light $\gg$ being a pedestrian on a pedestrian crossing $\gg$ originally being on the same road (as the competitor) while either going straight on or being (in) a tram $\gg$ having a diamond-shaped priority sign and not being a pedestrian $\gg$ being (in) a tram $\gg$ being (in) a car $\gg$ coming from the right $\gg$ being (on) a bicycle). EDCD would work perfectly if fed with traffic situations, and information as to 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 2002), the modelling of learnability in more comprehensive views of the grammar (e.g. Escudero & Boersma 2003), the modelling of covert subset phenomena (e.g. Prince & Tesar 1999 and Hayes, to appear), the modelling of optionality (e.g. Boersma & Hayes 2001) and the modelling of language change (e.g. Jäger 2003). For the general OT phonologist, the failures of the learning algorithms noted in this book and those 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 places large restrictions on what types of languages are possible and what types are not.

REFERENCES

Anttila, Arto (1997). *Variation in Finnish phonology and morphology*. PhD dissertation, 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**. 101–148.

Boersma, Paul (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* **21**. 43–58.

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

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

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

Eisner, Jason (2000). Easy and hard constraint ranking in Optimality Theory: algorithms and complexity. In Jason Eisner, Lauri Karttunen & Alain Thériault (eds.) *Finite-state phonology: Proceedings of the 5th Workshop of the ACL Special Interest Group in Computational Phonology* (*SIGPHON*). Luxembourg. 22–33.

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*. Philadelphia: University of Pennsylvania. 71–85.

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: Cambridge University Press.
Jäger, Gerhard (2003). Learning constraint sub-hierarchies: the Bidirectional Gradual Learning Algorithm. In Henk Zeevat & Reinhard Blutner (eds.) *Optimality Theory and pragmatics*. Basingstoke: Palgrave Macmillan. 251–287.
Lightfoot, David (1999). *The development of language: acquisition, change, and evolution*. Malden, Mass.: Blackwell.
Lombardi, Linda (1999). Positional faithfulness and voicing assimilation in Optimality Theory. *NLLT* **17**. 267–302.
McCarthy, John & Alan Prince (1995). Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*. Amherst: GLSA. 249–384.
Prince, Alan & Bruce Tesar (1999). *Learning phonotactic distributions*. Technical Report TR-54, Rutgers University Center for Cognitive Science.
Tesar, Bruce (1995). *Computational Optimality Theory*. PhD dissertation, University of Colorado, Boulder.
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*. Somerville, Mass.: Cascadilla Press. 615–626.
Tesar, Bruce (1998). An iterative strategy for language learning. *Lingua* **104**. 131–145.
Tesar, Bruce (1999). Robust interpretive parsing in metrical stress theory. *WCCFL* **17**. 625–639.
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*. Oxford: Oxford University Press. 592–620.
Tesar, Bruce & Paul Smolensky (1993). *The learnability of Optimality Theory: an algorithm and some basic complexity results*. Ms, University of Colorado, Boulder.
Tesar, Bruce & Paul Smolensky (1996). *Learnability in Optimality Theory* (*long version*). Technical Report 96:3, Department of Cognitive Science, Johns Hopkins University.
Tesar, Bruce & Paul Smolensky (1998). Learnability in Optimality Theory. *LI* **29**. 229–268.