

A Theory of Individual-Level Predicates Based on Blind Mandatory Implicatures. Constraint Promotion for Optimality Theory

by
Giorgio Magri

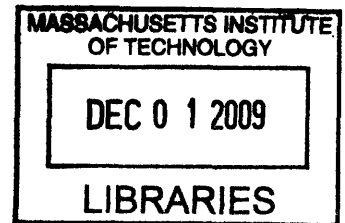
ARCHIVES

M.A. Philosophy, Università degli Studi di Milano, Italy, 2001
M.A. Mathematics, Università degli Studi di Milano, Italy, 2002

Submitted to the Department of Linguistics and Philosophy
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the
Massachusetts Institute of Technology
September 2009



©2009 Giorgio Magri. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Author
Department of Linguistics and Philosophy

Certified by
Adam Albright
Associate Professor of Linguistics
Thesis supervisor

Certified by
Danny Fox
Professor of Linguistics
Thesis supervisor

Accepted by
Irene Heim
Professor of Linguistics
Chair, Department of Linguistics and Philosophy

**A Theory of Individual-Level Predicates
Based on Blind Mandatory Implicatures.
Constraint Promotion for Optimality Theory**

by Giorgio Magri

Submitted to the Department of Linguistics and Philosophy on August 28, 2009,
in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Part I of this dissertation proposes an implicature-based theory of individual-level predicates. The idea is that we cannot say ‘#John is sometimes tall’ because the sentence triggers the scalar implicature that the alternative ‘John is always tall’ is false and this implicature mismatches with the piece of common knowledge that tallness is a permanent property. *Chapter 1* presents the idea informally. This idea faces two challenges. First, this scalar implicature must be mandatory and furthermore blind to common knowledge. Second, it is not clear how this idea extends to other properties of individual-level predicates. *Chapter 2* makes sense of the surprising nature of these special mismatching implicatures within the recent grammatical framework for scalar implicatures of Chierchia (2004) and Fox (2007a). *Chapter 3* shows how this implicature-based account can be extended to other properties of individual-level predicates, such as restrictions on their bare plural subjects, on German word order and extraction, and on Q-adverbs.

Part II of this dissertation develops a theory of update rules for the OT on-line algorithm that perform constraint promotion too, besides demotion. *Chapter 4* explains why we need constraint promotion, by arguing that demotion-only update rules are unable to model Hayes’ (2004) early stage of the acquisition of phonology. *Chapter 5* shows how to get constraint promotion, by means of two different techniques. One technique shares the combinatoric flavor of Tesar and Smolensky’s analysis of demotion-only update rules. The other technique consists of adapting to OT results from the theory of on-line algorithms for linear classification. The latter technique has various consequences interesting on their own, explored in *Chapter 8*. *Chapters 6* and *7* start the investigation of the properties of update rules that perform promotion too, concentrating on the characterization of the final vector and on the number of updates.

Thesis supervisors:	Adam Albright,	Danny Fox
Title:	Associate Professor of Linguistics	Professor of Linguistics

Academic acknowledgements

I do not like to say my things in public. Even less to write them down. So I will keep my written public academic acknowledgements short, and thank people in person. I wish to thank Adam Albright, Gennaro Chierchia, Danny Fox, and Irene Heim. I also wish to thank Marta Abrusan, Asaf Bachrach, Emmanuel Chemla, Cleo Condoravdi, Kai von Fintel, Mark Johnson, Angelika Kratzer, Sabine Iatridou, Livio Pizzocchero, Philippe Schlenker, Benjamin Spector, and Donca Steriade.

all.

Contents

Part I: A Theory of Individual Level Predicates Based on Blind and Mandatory Scalar Implicatures	13
1 Introduction	15
1.1 The Problem: restrictions on i-predicates	15
1.2 The idea: blind and mandatory mismatching implicatures	16
1.3 The solution: i-predicates and scalar implicatures	24
1.4 Implications	28
2 Oddness by mismatching scalar implicatures	29
2.1 The basic case	30
2.1.1 First step: covert ‘only’	31
2.1.2 Second step: blindness	32
2.1.3 Third step: mandatoriness	35
2.2 Extension to cases with multiple alternatives	41
2.3 Extension to downward entailing environments	50
2.4 Extension to Presuppositions	60
2.5 Miscellaneous issues	68
2.5.1 An alternative approach based on Manner and its inadequacy	68
2.5.2 Predecessors	70
2.5.3 Miscellaneous problematic cases	73
3 Application to individual level predicates	75
3.1 Existential Q-adverbs	77
3.1.1 Existing accounts	77
3.1.2 An account based on blind and mandatory mismatching implicatures	82
3.2 Temporal modification	86
3.2.1 Existing accounts	86
3.2.2 An account based on blind and mandatory mismatching implicatures	88
3.3 Bare plural subjects	93
3.3.1 Existing accounts	94
3.3.2 An account based on blind and mandatory mismatching implicatures	105
3.4 Word order and extraction in German	110
3.4.1 Existing accounts	111
3.4.2 An account based on blind and mandatory mismatching implicatures	112
3.5 BPSs embedded under a universal operator	116
3.5.1 Existing accounts	117
3.5.2 An account based on blind and mandatory mismatching implicatures	120
3.6 More facts on the distribution of existential BPSs	123
3.6.1 First set of facts: existential BPSs that are non-kind denoting	123
3.6.2 Second set of facts: BPSs associated with ‘only’	127
3.6.3 Third set of facts: generalized locatives?	128
3.6.4 Fourth set of facts: statives	129
3.7 Overt universal Q-adverbs	132

Part II: Constraint promotion and the OT on-line model for the acquisition of phonology	137
4 Why we need constraint promotion	139
4.1 OT preliminaries	140
4.1.1 OT basics	140
4.1.2 OT with comparative rows	142
4.1.3 OT with ranking vectors	145
4.2 The OT on-line algorithm: a computational perspective	148
4.2.1 Outline of the algorithm	148
4.2.2 Generalities on demotion-only update rules	153
4.2.3 First example of demotion-only update rule: minimal and gradual	154
4.2.3.1 Bound on the worst-case number of updates	155
4.2.3.2 Characterization of the final vector	157
4.2.3.3 Extension to an arbitrary initial vector	158
4.2.4 Second example of demotion-only update rule: minimal and non-gradual	160
4.2.5 Third example of demotion-only update rule: non-minimal and gradual	162
4.2.6 Promotion-demotion update rules	164
4.2.6.1 A detailed explanation of Pater's counterexample	165
4.2.6.2 On the analysis of promotion-demotion update rules	169
4.3 The OT on-line algorithm: a modeling perspective	172
4.3.1 Initialization	172
4.3.2 Choice of the underlying form	172
4.3.3 A toy example	173
4.3.4 How to choose update rules	175
5 How to get constraint promotion	179
5.1 A principled promotion-demotion update rule	180
5.2 First proof of finite time convergence	185
5.2.1 First step	185
5.2.2 Second step	186
5.2.3 Third step	187
5.2.4 Fourth step	191
5.3 Variants	191
5.3.1 First variant	191
5.3.2 Second variant	192
5.3.3 Third variant	193
5.4 Second proof of finite time convergence	194
5.4.1 First step	195
5.4.2 Second step	195
5.4.3 Third step	197
5.5 More variants	198
5.5.1 First variant	199
5.5.2 Second variant	200
5.5.3 Third variant	203
Appendix: some classical results on linear on-line algorithms	204
6 How to study constraint promotion: the final ranking vector	209
6.1 An invariant	210
6.2 Preliminaries on OT phonotactics	215
6.3 Test cases	220
6.3.1 First test case	221
6.3.2 Second test case: pseudo-korean	225
6.3.2.1 Stage I	228
6.3.2.2 Stage II	229

6.3.2.3 Stage III	232
6.3.2.4 Stage IV	233
6.3.3 Third test case: the whole Azba typology	233
7 How to study constraint promotion: number of updates	245
7.1 Constraint promotion might require many updates	246
7.1.1 A simple lower-bound on the best-case number of updates	247
7.1.2 A case with an exponential best-case number of updates	249
7.1.3 A small remark on Pater's comparative tableaux	249
7.2 Formulations of the problem of the acquisition of phonology	252
7.2.1 The Ranking problem	252
7.2.2 The Subset problem	254
7.2.3 Prince and Tesar's (2004) formulation of the Subset problem	255
7.3 How to decide whether a problem is easy or not	255
7.3.1 First step	256
7.3.2 Second step	257
7.3.3 Third step	257
7.3.4 Fourth step	258
7.4 The problem of the acquisition of phonology is "hard"	258
7.4.1 Complexity of the Ranking problem	259
7.4.2 Complexity of Prince and Tesar's Subset problem	261
7.4.3 Complexity of the Subset problem	267
8 Consequences for the relationship between standard and linear OT	271
8.1 The relationship between OT-compatibility and L-compatibility	272
8.2 Computational consequences	276
8.3 More consequences	279
List of symbols used in part II	281
References	293

Part I

A Theory of Individual-Level
Predicates Based on Blind Mandatory
Scalar Implicatures

Abstract of Part I — Predicates such as ‘tall’ or ‘to know Latin’, which intuitively denote permanent properties, are called INDIVIDUAL-LEVEL PREDICATES. Many peculiar properties of this class of predicates have been noted in the literature. One such property is that we cannot say ‘#John is sometimes tall’. Here is a way to account for this property: this sentence sounds odd because it triggers the scalar implicature that the alternative ‘John is always tall’ is false and this implicature mismatches with the piece of common knowledge that tallness is a permanent property. This intuition faces two challenges. *First*: this scalar implicature has a odd status, since it must be surprisingly mandatory (otherwise, it could be cancelled and the sentence rescued) and furthermore blind to common knowledge (since the piece of common knowledge that tallness is a permanent property makes the two alternatives equivalent). *Second*: it is not clear how this intuition could be extended to other properties of individual-level predicates. Part I of this dissertation defends the idea of an implicature-based theory of individual-level predicates by facing these two challenges. *Chapter 1* presents the idea informally. *Chapter 2* tries to make sense of the surprising nature of these special mismatching implicatures within the recent grammatical framework for scalar implicatures of Chierchia (2004) and Fox (2007a). *Chapter 3* shows how this implicature-based line of reasoning can be extended to other properties of individual-level predicates, such as restrictions on the interpretation of their bare plural subjects, noted in Carlson (1977), Milsark (1977), and Fox (1995); restrictions on German word order and extraction, noted in Diesing (1992); and restrictions on Q-adverbs, noted in Kratzer (1995).

Chapter 1

Introduction

1.1 The Problem: restrictions on i-predicates

Predicates such as those in (1a) intuitively denote permanent properties, while those in (1b) denote properties *not* necessarily permanent. Milsark (1974) dubs possibly temporary predicates like ‘available’ in (1b) as STATE DESCRIPTIVE predicates and persistent or permanent predicates like ‘tall’ in (1a) as PROPERTY predicates; Carlson (1977) calls them STAGE LEVEL predicates (henceforth: s-predicates) and INDIVIDUAL LEVEL predicates (henceforth: i-predicates), respectively; I will stick to Carlson’s terminology, because it has had more success in the subsequent literature.

- (1) a. to be tall, to be related to Chomsky, to know Latin, ...
- b. to be available, to talk to Chomsky, to study Latin, ...

An impressive list of linguistic facts have been pointed out that set the two classes of predicates apart. The overall picture is that s-predicates can do many more things than i-predicates can do: there are configurations where s-predicates are judged fine but i-predicates are not (e.g. adverbial quantification, ‘there’-insertion, perception sentences, etc.) and there are readings that are judged available with s-predicates but not with i-predicates (e.g. the existential reading for their BPSs, the episodic reading, etc.). To illustrate, I review a few of these facts in the rest of this section; see Fernald (2000) for a comprehensive review. A theory of i-predicates should thus answer the following question: why is it the case that i-predicates cannot do the many things that s-predicates can do? This is the question addressed in part I of my dissertation.

Adverbs Contrary to s-predicates, i-predicates do not admit adverbial quantification of any quantificational force, as shown in (2) and (3).

- (2) a. John is sometimes available.
- b. #John is sometimes tall.
- (3) a. John is always available.
- b. #John is always tall.

Kratzer (1995) adds the observation that adverbial quantification with i-predicates turns fine in the presence of bare plurals or indefinites, as shown in (4).

- (4) a. Firemen are sometimes / always tall.
- b. A fireman is sometimes / always tall.

Modification Contrary to s-predicates, i-predicates do not allow temporal modifiers such as ‘after dinner’, as shown in (5b).

- (5) a. John is available after dinner.

- b. #John is tall after dinner.

Kratzer (1995) and Musan (1997) discuss the case of temporal modification through tense morphology. They note that sentence (6b), with the *i*-predicate ‘tall’ in the past tense, triggers the inference that John is dead, while sentence (6a), with the *s*-predicate ‘available’, does not. Following Musan (1997), I will refer to this inference as the LIFE-TIME EFFECT.

- (6) a. John was available. no life-time effect
 b. John was tall. life-time effect

Bare plural subjects Milsark (1977) and Carlson (1977) point out the contrast in (7): the bare plural subject (henceforth: BPS) ‘firemen’ of the *s*-predicate ‘available’ in (7a) admits both the generic reading (“Firemen are usually available people”) and the existential reading (“There are firemen who are available”); the BPS of the *i*-predicate ‘tall’ in (7b) instead lacks the existential reading and only allows the generic one.

- (7) a. Firemen are available. \exists -BPS, GEN-BPS
 b. Firemen are tall. * \exists -BPS, GEN-BPS

Fox (1995) adds the surprising observation that this restriction on the readings of BPSs of *i*-predicates is waived when BPSs are embedded under a universal quantifier: as expected, the BPS ‘Jewish women’ of the *i*-predicate ‘to be related to Chomsky’ of sentence (8a) lacks the existential reading; in sentence (8b), the definite ‘Chomsky’ has been replaced by the universal quantifier ‘every Jewish man’ and the BPS ‘Jewish women’ of sentence (8b) does admit the existential reading, provided that the universal object is assigned wide scope (“For every Jewish man *there are* Jewish women related to him”).

- (8) a. Jewish women are related to Chomsky. * \exists -BPS
 b. Jewish women are related to every Jewish man. \exists -BPS

German word order Diesing (1992) points out the contrast in (9): the BPS ‘Feuerwehrmänner’ (‘firemen’) can sit both at the left and at the right of the particles ‘ja doch’ (‘indeed’) in the case of the *s*-predicate ‘verfügbar’ (‘available’), as shown in (9a) and (9b); but it can sit only at the left of ‘ja doch’ in the case of the *i*-predicate ‘intelligent’, as shown by the contrast between (9c) and (9d).

- (9) a. ... weil ja doch Feuerwehrmänner verfügbar sind.
 ... since PARTs firemen available are
 b. ... weil Feuerwehrmänner ja doch verfügbar sind.
 c. ... *weil ja doch Feuerwehrmänner intelligent sind.
 ... since PARTs firemen intelligent are
 d. ... weil Feuerwehrmänner ja doch intelligent sind.

1.2 The idea: blind and mandatory mismatching implicatures

The solution to this problem which I will defend in this work hinges on the idea of blind and mandatory scalar implicatures that mismatch with common knowledge. This idea was first introduced in the literature by Hawkins (1978, 1991) and criticized by Heim (1991), Percus (2006) and Sauerland (2008). In this section, I introduce the idea in a nutshell, by reconstructing its history.

Hawkins’ puzzle Hawkins noted the oddness of sentences such as those in (10).¹ Intuitively, the oddness of these sentences is due to the fact that we know that the victim had a unique father, that our tent has a unique weight, and so on. In all these cases where the restrictor is known to denote a singleton, we cannot use the indefinite ‘a’ but have to use the singular definite ‘the’ instead.

¹Naively assuming that there is a specific judgment of “oddness”, I use the diacritic ‘#’ to signal it.

- (10) a. #John has interviewed a father of the victim.
 b. #A weight of our tent is under four lbs.
 c. #Fred lost a nose in the war.
 d. #I didn't buy the house because a roof was leaking.
 e. #I peered into a center of the flower.
 f. #I climbed to a top of the tree.

Consider for instance Hawkins' sentence (10a), repeated below as (11a), together with its fine variant (11b). Assume that 'a' is a standard existential quantifier. Then, there is no contradiction between sentence (11a) and the common knowledge that the victim had a unique father. Thus, we might have expected that an utterance of sentence (11a) would have been "additively integrated" with the common knowledge that people have only one father, thus boiling down to the statement that John has interviewed *the* father of the victim, precisely as stated by sentence (11b). But that doesn't happen.

- (11) a. #John has interviewed a father of the victim.
 b. John has interviewed the father of the victim.

The oddness of Hawkins' sentences (10) thus seems to suggest that these sentences somehow trigger an ANTIUNIQUENESS INFERENCE that the restrictor NP is not a singleton or that it is not known to be a singleton. Once this inference is in place, we can straightforwardly account for the oddness of Hawkins' sentences (10) by means of the piece of reasoning in (12).

- (12) a. On the one hand, Hawkins' sentences (10) somehow trigger an antiuniqueness inference.
 b. On the other hand, common knowledge ensures that uniqueness holds.
 c. In conclusion, the oddness of Hawkins' sentences (10) follows from the mismatch between the inference in (12a) and common knowledge in (12b).

But what is the nature of this antiuniqueness inference? in other words, how can the first step (12a) of the preceding account be made explicit? Let me dub this question as HAWKINS' PUZZLE. To answer it, we need to go through our typology of inferences, and decide which one fits best the case of the antiuniqueness inference triggered by sentences (10). Let me consider three different options in turn.

The naïve solution: antiuniqueness as a presupposition Sentences (13a) and (14a) trigger the inference that the corresponding sentences (13b) and (14b) are common knowledge. These inferences are called PRESUPPOSITIONS.

- (13) a. The father of the victim arrived late.
 b. The victim has one and only one father.
 (14) a. John knows that this solution is wrong.
 b. This solution is wrong.

A rather straightforward solution to Hawkins' puzzle is to assume that the anti-uniqueness inference triggered by sentences (10) is a presupposition. In other words, the standard semantics for 'a' as a bare existential quantifier is wrong, and should be replaced with the presuppositional variant (15), according to which 'a' presupposes that its restrictor is not a singleton. Let me call ANTIUNIQUENESS PRESUPPOSITION the presupposition carried by 'a' according to (15). Note that this presupposition is the negation of the presupposition usually associated with the singular definite article.

$$(15) \llbracket a \rrbracket = \lambda P . \lambda Q : |P| > 2 . P \cap Q \neq \emptyset$$

Under this assumption (15) that antiuniqueness is a presupposition, the general explanation scheme (12) can be made explicit as the account in (16).

- (16) a. On the one hand, Hawkins' sentences (10) trigger the antiuniqueness *presupposition*.

- b. On the other hand, common knowledge ensures that uniqueness holds.
- c. In conclusion, the oddness of Hawkins' sentences (10) follows from the mismatch between the presupposition in (16a) and common knowledge in (16b).

In other words, (16) says that the oddness of Hawkins' sentences (10) is an instance of the very well known phenomenon of presupposition failure.

Against the naïve solution Plausible as it might look at first sight, this solution to Hawkins' puzzle turns out to be on the wrong track, as shown by the following three arguments. A *first* argument, due to Heim (1991, §2.1.3), is based on sentences (17). Heim notes that sentence (17a) does not presuppose there to be at least two 20 ft. long catfishes: "whoever asserts or hears [(17a)] may very well assume that Robert was lucky enough to catch by far the longest catfish in the world"; nor does sentence (17b) presuppose that the speaker has two pathologically nosy neighbors: the sentence "leaves quite open how many pathologically nosy neighbors I have and in no way discourages the hope that it's only one." Of course, these intuitions about sentences (17) are incompatible with the hypothesis that 'a' bears the antiuniqueness presupposition (15).

- (17) a. Robert caught a 20 ft. long catfish.
- b. A pathologically nosy neighbor of mine broke into the attic.

A *second* argument, due to Sauerland (2003c, 2008), is based on the contrast in (18). As Sauerland points out, sentence (18a) suffers from presupposition failure, since the uniqueness presupposition of the singular definite article 'the' projects universally (under plausible assumptions on presupposition projection under 'every'; see for instance Heim (1988) for discussion) and it thus fails in the scenario considered, where some applicants have written more than one paper. If 'a' had an antiuniqueness presupposition as in (15), then sentence (18b) would be incorrectly predicted to suffer from presupposition failure too, since this presupposition would be expected to project universally as well and thus to fail in the scenario considered, where some applicants have written only one paper.

- (18) *Context:* several candidates applied. Some have written only one paper, others have written more than one. The selection committee decides:
 - a. #Every candidate should send his paper.
 - b. Every candidate should send a paper of his.

Let me add a *third* argument on top of these two, based on sentence (19). Intuitively, this sentence sounds odd because it mismatches with the piece of common knowledge that people have two eyes of the same color. Crucially, the antiuniqueness presupposition (15) of 'a' would be trivially satisfied in the case of sentence (19) and thus cannot play any role in accounting for its oddness. This suggests that something else is needed, also in the case of Hawkins' sentences (10) we started with.

- (19) #An eye of the victim is blue.

In conclusion, adding the antiuniqueness presupposition (15) to the semantics of 'a' doesn't quite seem right, because of Heim's argument (17) and Sauerland's argument (18); nor useful, because it misses the generalization that (19) feels odd in the same way that Hawkins' sentences (10) do. Let's thus abandon (15) and stick with the more traditional semantics for 'a', according to which it is a plain existential quantifier. We are thus back to square one: what is the nature of the uniqueness inference triggered by Hawkins' sentences (10)?

Hawkins' solution: antiuniqueness as a scalar implicature Sentences (20a) and (21a) trigger the inference that the corresponding sentences (20b) and (21b) are false. These inferences are called SCALAR IMPLICATURES.

- (20) a. John did some of the homework.
- b. John did all of the homework.
- (21) a. John met Mary or Sue.

- b. John met Mary and Sue.

Hawkins's suggests that the antiuniqueness inference triggered by sentences (10) is a scalar implicature. Thus, the general explanation scheme (12) becomes the account sketched in (22).

- (22) a. On the one hand, sentence (11a) triggers the scalar implicature that the speaker is not in a position to utter (11b), i.e. does not know whether the victim had a unique father.
 b. On the other hand, common knowledge entails that people have a unique father, namely that (11a) cannot be true while (11b) false.
 c. In conclusion, the oddness of (11a) follows from the mismatch between the scalar implicature in (22a) and the common knowledge in (22b).

The idea sketched in (22) might cope well with the last two challenges for the presuppositional account (15) discussed above. Sauerland's problem in (18) has obviously disappeared, since now we can assume the standard nonpresuppositional semantics for 'a'. Of course, Hawkins' reasoning (22) predicts that Sauerland's sentence (18b) triggers the antiuniqueness inference that it is false that every candidate wrote a unique paper and thus sentence (18b) should sound odd in a context where every candidate wrote a unique paper. This prediction is right. To see that more clearly, let me switch to the formally analogous example (23b). Sentence (23a) is from Roberts (2003), who in turns takes it from Kadmon (1987). Sentence (23b) is the minimal variant with the definite replaced by the indefinite. It seems to me that (23b) sounds odd, namely it mismatches with the piece of common knowledge that *all* unicycles have a single wheel.

- (23) a. Every unicycle had a spoke missing from *the* wheel.
 b. #Every unicycle had a spoke missing from *a* wheel.

Finally, we might reasonably hope that this account (22) can be extended to the case of the odd sentence (19) too.

Against Hawkins' solution According to Grice (1975), scalar implicatures are rational pragmatic inferences triggered by the Maxim of Quantity (24). Let me quickly review the core idea of the proposal with the classical cases (20) and (21): sentences (20b) and (21b) are stronger than sentences (20a) and (21a), in the sense that the proposition denoted by the former is a subset of the proposition denoted by the latter; thus (20b) and (21b) are more informative than (20a) and (21a), under any plausible construal of informativeness; by the maxim of Quantity (24), the speaker would thus have used these more informative alternatives, had he judged them true; since he didn't use then, then he must have judged them false.

- (24) Among a set of alternatives, use the most informative sentence you believe to be true.

The first step (22a) of Hawkins's account says that the antiuniqueness inference triggered by sentences (10) is a scalar implicature. To make this step of the account compatible with Grice's account of scalar implicatures, Hawkins assumes a Russellian non-presuppositional semantics for the definite article, whereby a sentence containing a singular definite *entails* existence and uniqueness. Thus, a sentence with a singular definite is usually stronger than the corresponding sentence with an indefinite. Yet, this move is by far not sufficient to make Hawkins' proposal compatible with Grice's theory of scalar implicatures. Heim (1991, §2.1.3) notes that the first step (22a) of Hawkins's account is incompatible with Grice's theory of scalar implicatures because of the issue summarized in (25).² I will call (25) the BLINDNESS ISSUE: why should the Maxim of Quantity be blind to common knowledge? I will sharpen the issue further in subsection ??.

²Here is Heim's passage: "The standard examples of scalar implicatures are usually derived from the Gricean conversational Maxim of Quantity ('Make your contribution as informative as is necessary given the purpose of the conversation!'). At this point, however, our analogy breaks down, since no such derivation seems possible for the non-uniqueness implicature of the indefinite article. If it is already known that each person has only one father, [(11a)] conveys exactly the same amount of new information as the corresponding sentence with the definite article [(11b)]. Lack of informativeness can therefore not be the reason for the inadequacy here".

- (25) **BLINDNESS ISSUE.** Within the Gricean theory, scalar implicatures are triggered by a violation of the maxim of Quantity. Being a pragmatic maxim, Quantity compares an utterance to its alternatives with respect to their informativeness. But because of common knowledge, utterances of (11a) and (11b) convey exactly the same information. Thus no violation of Gricean Quantity can arise by uttering (11a) instead of (11b).

Another issue that makes Hawkins' reasoning (22) hardly compatible with Grice's theory of scalar implicatures is (26). I will call (26) the **MANDATORINESS ISSUE**: why should these mismatching implicatures be mandatorily derived? Heim's problem in (17) illustrates this issue: why is it the case that the indefinite in sentences (17) does not trigger the anti-uniqueness scalar implicature that sentences (10) trigger?

- (26) **MANDATORINESS ISSUE.** Within the Gricean theory, scalar implicatures are pragmatic inferences. Hence, they have a weak status: they are optional, cancellable, and suspendable. Thus, it is not at all clear why the mismatching implicature is kept in place and an utterance of (11a) deemed odd, rather than the implicature cancelled or suspended or never computed, and thus the utterance rescued.

Grice's original approach has been developed, formalized, and defended by a number of authors, such as Horn (1972), Gazdar (1979), Sauerland (2004c), Spector (2007c), Horn (2005), Russell (2006), and Geurts (2009), among others. These various neo-Gricean implementations all share the core idea of a *pragmatic* derivation of scalar implicatures: grammar pairs each sentence with a *plain meaning* and scalar implicatures are then computed on the basis of this meaning by an extragrammatical procedure rooted in general-purpose principles of rationality. As far as I can see, the two issues laid out in (25) and (26), which arise when Hawkins' suggestion (22) is implemented within Grice's theory, also arise within any other pragmatic theory of scalar implicatures.³ Let me state this point explicitly, as in (27) and (28). I conclude that Hawkins's account (22) is intrinsically incompatible with pragmatic approaches to scalar implicatures.

- (27) **GENERALIZED BLINDNESS ISSUE.** If implicatures are computed by a pragmatic engine, how can this engine be blind to common knowledge so as to distinguish between alternatives that are equivalent given common knowledge?
- (28) **GENERALIZED MANDATORINESS ISSUE.** If implicatures are computed by a pragmatic engine, how can this engine be robust enough to mandatorily force the mismatching implicature in place?

Heim's reaction to the incompatibility between Grice's theory of implicatures and Hawkins' reasoning (22) was to reject the latter, and to sketch an alternative account for the oddness of sentences (10), which has nothing to do with scalar implicatures.

Heim's solution: antiuniqueness as an anti-presupposition Sentence (29a) triggers the inference that the corresponding sentence (29b) is false or at least not taken for granted. Following Percus (2006) and Chemla (2008), let me call inferences such as the one illustrated in (29) **ANTI-PRESUPPOSITIONS**.⁴

- (29) a. John believes that Mary is pregnant.
b. Mary is pregnant.

Sauerland (2008) argues that anti-presuppositions are not presuppositions by means of the contrast in (30), analogous to the one in (17) considered above. The oddness of the continuation in (30a)

³As pointed out to me by A. Kratzer (p.c.), this claim is not quite accurate for the case of Gazdar (1979). His system dispenses altogether with Grice's Maxims, so that the current statement (25) is moot in the case of Gazdar. Yet, his definition of "context update" (pp. 131-132; definitions XVI-XVII) is unable to derive the mismatching implicature needed by Hawkins to account for the oddness of sentence (11a). Furthermore, Gazdar's system does warrant a certain degree of mandatoriness to implicatures, since he assumes that when an implicature and a presupposition mismatch, it is the implicature that wins over the presupposition rather than the implicature being cancelled by the presupposition (see also subsection 2.5.2 below).

⁴Sauerland (2008) calls these inferences **IMPLICATED PRESUPPOSITIONS**.

shows that the factivity presupposition triggered by ‘know’ projects universally from underneath the universal quantifier ‘every’. If the anti-factivity inference of ‘believe’ in (29a) were a presupposition, then it would be expected to project universally too. But this prediction is contradicted by the fact that the continuation in (30b) is perfectly fine.⁵

- (30) a. Every audience member knows that his support was crucial for the team.
 ... #But only John’s support was indeed crucial
- b. Every audience member believes that his support was crucial for the team.
 ... But only John’s support was indeed crucial

If anti-presuppositions cannot be encoded directly into the semantics as presuppositions, they must come about in some more indirect way. There seems to be agreement in the literature that anti-presuppositions come about through some sort of “competition” with corresponding presuppositional alternatives. For example, that the anti-factivity inference of sentence (29a) somehow comes about through “competition” with the alternative (31) whose predicate ‘know’ carries a factivity presupposition. But what is exactly the nature of this competition?

- (31) John knows that Mary is pregnant.

At this point, anti-presuppositions start to look very much like scalar implicatures. Yet, it has been suggested that they are not scalar implicatures. On the one hand, the hypothesis that the anti-factivity presupposition of sentence (29a) comes about as a scalar implicature triggered by competition with the presuppositional alternative (31) runs into the issue (32), which is reminiscent of the Blindness issue noted above in (25) or (27) with the first step of Hawkins’ reasoning (22).⁶

- (32) If the factivity presupposition carried by (31) is satisfied, then it is already common knowledge that Mary is pregnant and thus it is not clear why (31) should count as more informative than (29a) and thus preferable from the point of view of the maxim of Quantity (24). If the factivity presupposition carried by (31) is not satisfied, then the alternative (31) would be infelicitous and it could thus hardly compete with (29a).

On the other hand, Sauerland (2008) provides some empirical arguments that anti-presuppositions behave differently from scalar implicatures. For instance, he notes that “a [...] contrast between implicatures and [anti-presuppositions] is their projection through negation. [Anti-presuppositions] just like conventional presuppositions are not affected by negation. [Anti-presuppositions] clearly contrast with scalar implicature in this way. Scalar implicatures are reversed in the scope of negation and other downward entailing operators.” To illustrate this special behavior of anti-presuppositions, Sauerland notes that the anti-factivity anti-presupposition of sentence (29a) does not disappear when ‘believe’ is embedded in a DE environment, as in (33).

- (33) I doubt that John believes that Mary is pregnant.

⁵Another way of checking whether anti-presuppositions are presuppositions or not is to use the “Hey, wait a minute” test, devised in von Stechow (2003) building on Shanon (1976). As shown in (i), true presuppositions pass the test. My judgment concerning (ii) is not sharp. But I am inclined to say that the “Hey wait a minute” continuation in (ii) is worse than in (i). If this judgment is correct, then the “Hey, wait a minute” test too shows that anti-presuppositions do not pattern with presuppositions.

- (i) A: John knows that Mary is pregnant.
 B: Hey wait a minute: I didn’t know that Mary is pregnant.
- (ii) A: John believes that Mary is pregnant.
 B: #Hey wait a minute, I thought that we knew whether she is or not.

Further evidence that anti-presuppositions are not presuppositions is provided in Yatsushiro (2005). Following for instance Sauerland (2008), she assumes that the existence inference triggered by ‘every’ is a presupposition while the anti-uniqueness inference triggered by ‘every’ is an anti-presupposition. She reports the following finding on 16 English speaking children in the age range 3:11-5;11: “The subjects responded adult-like 72% of the time [...] for the items testing for the knowledge of the existence presupposition, whereas they responded 32% of the time [...] for the anti-uniqueness presupposition.”

⁶Schlenker (2006) offers a detailed discussion of this issue (32), based on the refined notion of common ground of Stalnaker (2002).

Since anti-presuppositions are neither presuppositions nor scalar implicatures, they must be a third class of inferences, derived by an independent, dedicated mechanism. The core of this mechanism is the idea that the use of presuppositional items is tightly constrained by dedicated principles, such as the one loosely stated in (34) after Heim (1991). Following Sauerland (2008), I refer to (34) as MAXIMIZE PRESUPPOSITION (henceforth: MP). The anti-factivity presupposition triggered by (29a) thus arises by MP-competition with (31). MP is, in a sense, a variant of the maxim of Quantity (24) that applies at the level of presuppositions. Yet, the two principles crucially differ in nature: the maxim of Quantity (24) can be construed as a pragmatic, extra-grammatical principle that follows from general principles of rationality; because of the issue in (32), MP must instead be construed as a principle of grammar.

(34) Among a set of alternatives, use the felicitous sentence with the strongest presupposition.

Having enriched our typology of inferences with anti-presuppositions, we can now go back to Hawkins' puzzle. If we assume the standard presuppositional semantics for the definite article 'the', then it is natural to construe the anti-uniqueness inference triggered by the indefinite 'a' in Hawkins' sentences (10) as an anti-presupposition triggered by MP-competition with the singular definite article. Under this assumption, the general scheme (12) becomes the account in (35).

- (35) a. On the one hand, sentence (11a) triggers the anti-presupposition that it does not follow from common knowledge that the victim had a unique father.
 b. On the other hand, common knowledge entails that people have a unique father.
 c. In conclusion, the oddness of (11a) follows from the mismatch between the anti-presupposition in (35a) and the common knowledge in (35b).

The conclusion of the existing literature, with the partial exception of Schlenker (2006), seems to be that anti-presuppositions are a separate, special class of inferences; that they are derived by a dedicated grammatical constraint such as MP (34); and that the proper account of Hawkins' sentences is in terms of anti-presuppositions, as in (35). See Sauerland (2008), Percus (2006) and Chemla (2008) for refinements of this line of reasoning. And for other applications besides Hawkins' sentences (10), see for instance Sauerland (2002, 2003a,c, 2004a), Schlenker (2005), Amsili and Beyssade (2006), among others.

Against Heim's solution In a (p.c.) reported in Schlenker (2006), E. Chemla makes the point that MP is too narrow in coverage because it does not extend to cases of oddness intuitively similar to (10), where nonetheless presuppositions plausibly do not play any role. Let me discuss a slight variant of Chemla's argument.⁷ Consider sentence (36a), next to one of Hawkins' sentences (10), such as the one repeated in (36b). Sentence (36a) is a small variant of sentence (36b), in which the indefinite restricts a plural noun instead of a singular one. Both sentences sound odd: sentence (36a) mismatches with the piece of common knowledge that people have only two parents; sentence (36b) mismatches with the piece of common knowledge that people have only one father. Let me thus investigate this small extension of Hawkins' puzzle.

- (36) a. #Some parents of the victim got married (to each other) in the spring of 1972.

⁷Chemla original argument was based on (i).

- (i) John assigned the same grade to all of his students...
 #He gave an A to *some* of them.

I cannot use this example (i) to argue against MP. In fact, in section 2.4 I will have to resort to the assumption that plural definites trigger a HOMOGENEITY PRESUPPOSITION. Thus, sentence (ii) carries the presupposition that either John gave an A to all of his students or he gave an A to none of them.

- (ii) He gave them an A.

Thus, sentence (ii) does carry a presupposition that the original sentence (i) does not carry, and therefore MP (34) does account for the oddness of the original sentence (i). To overcome this slight technical problem, I have replaced Chemla's example (i) with the variant (36a), where the homogeneity presupposition is neutralized by the fact that the predicate 'to get married (to each other)' is collective.

- b. #John interviewed a father of the victim.

Hawkins' accounts for the oddness of sentence (36b) by competition with the corresponding sentence with the indefinite replaced by the definite. Hawkins's reasoning straightforwardly extends to the case of (36a), as made explicit in (37).

- (37) a. On the one hand, sentence (36a) triggers the scalar implicature that the speaker is not in a position to utter the corresponding sentence with the indefinite replaced by the definite, i.e. the speaker does not know whether all parents of the victim got married in 1972.
 b. On the other hand, common knowledge entails that people have only two parents, namely that the truth of sentence (36a) entails the truth of the corresponding sentence with the indefinite replaced by the definite.
 c. In conclusion, the oddness of sentence (36a) follows from the mismatch between the scalar implicature in (37a) and the common knowledge in (37b).

Heim accounts for the oddness of sentence (36b) by MP-competition with the corresponding sentence with the indefinite replaced by the definite. Let me argue that Heim's account does not extend to (36a), and is thus too narrow in coverage. As far as I can see, there are three relevant alternatives for MP-competition with sentence (36a) to be considered, listed in (38).⁸

- (38) a. *The* parents of the victim got married in the spring of 1972.
 b. *The two* parents of the victim got married in the spring of 1972.
 c. *?Both* parents of the victim got married in the spring of 1972.

For the case (36b) of the singular noun, I have considered the alternative obtained by replacing the indefinite with the definite and let MP make leverage on the *uniqueness* presupposition triggered by the latter. But in the case of plural nouns, the alternative (38a) with the definite does not do the work, since there is no uniqueness presupposition in the case of plural nouns. One might still invoke MP with respect to the *existential* presupposition of the definite in (38a). But this strategy can hardly be on the right track. Consider the variant of sentence (36a) in (39), obtained by replacing the plain indefinite with a partitive. This variant (39) sounds just as odd as the sentence (36a) that I started with. Yet, both (39) and the alternative (38a) now bear the existential presupposition, and thus MP really has nothing to work on. In any case, the existential presupposition of the alternative (38a) doesn't really seem to be playing any role here: the intuitive reason why sentence (36a) sounds odd is that it mismatches with the common knowledge that people have two parents, not with the common knowledge that people do have parents.

- (39) #Some of the parents of the victim got married in the spring of 1972.

This failure with the alternative (38a) might suggest to slightly modify this initial alternative (38a) as in (38b). If indeed (38b) were an MP-alternative of (36a), then the oddness of the latter could be accounted for by the fact that the alternative (38b) carries the presupposition that the victim had exactly two parents. But I think that this strategy does not work, because it is implausible that (38b) is an MP-alternative of (36a). In fact, assume by contradiction that (38b) were indeed an MP-alternative of (36a). If 'some' and 'the two' are alternatives, then *a fortiori* should 'the' and 'the two' be alternatives as well, since they are even "closer to each other", so to speak. But the latter cannot be MP-alternatives. In fact, if 'the' and 'the *n*' were MP-alternatives, then the fine sentence (40) would be incorrectly ruled out by MP, as noted by Chemla (2006).

- (40) Philippe broke the fingers of his left hand.

Finally, I submit that the alternative (38c) does not do the work simply because 'both' forces a distributive reading which is incompatible with collective predicates, such as 'got married (to each other) in 1972'.

⁸Here and throughout this work, italics in the examples is used to highlight to the reader some special features of the example, not to signal focus. In order to signal focus, I use the subscript 'F'.

Back to Hawkins' proposal The main reason for abandoning Hawkins' reasoning (22) had to do with its incompatibility with a theory whereby scalar implicatures are construed as pragmatic, extra-grammatical inferences, as stated in (25)-(26) and (27)-(28). Yet, the pragmatic theory of implicatures has recently been challenged for independent reasons and an alternative *grammatical* theory of scalar implicatures has been suggested, in works such as Chierchia (2004), Fox (2007a), Chierchia (2006a), and Chierchia et al. (to appear), among others. According to this alternative theory, scalar implicatures are derived through a purely grammatical algorithm. For instance, according to Fox's (2007) implementation of this idea, scalar implicatures are brought about by a covert version of 'only' syntactically realized at LF. The goal of chapter 2 is to argue in favor of Hawkins' account (22) from the perspective of this new grammatical theory of implicatures. I will try to put forward a variant of the grammatical theory of implicatures that copes with the Blindness issue (25)-(26) and the Mandatoriness issue (27)-(28). Furthermore, I will argue that, once the Blindness issue is taken care of, the related issue in (32), against construing anti-presuppositions as plain scalar inferences, is neutralized. Finally, I will argue that Sauerland's empirical arguments that scalar implicatures and anti-presuppositions have different empirical properties, such as the one in (33), does not hold against closer scrutiny. I will thus suggest that anti-presuppositions are not a special, separate class of inferences but run-of-the-mill scalar implicatures, with the only special but irrelevant property that they are triggered by presuppositional alternatives.

1.3 The solution: i-predicates and scalar implicatures

The version of the theory of scalar implicatures developed in chapter 2 provides the background for a theory of i-predicates based on blind and mandatory scalar implicatures that mismatch with what common knowledge encodes about i-predicates. Predecessors of this idea are Musan (1997), Maienborn (2004) and Percus (1997, Chapter 2). I sketch the idea in a nutshell in this section.

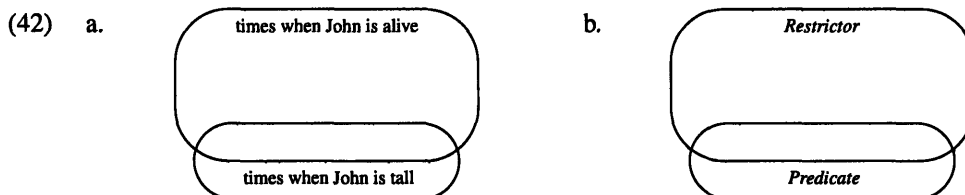
The classical solution As reviewed in Subsection 1.1, the s-/i-predicates distinction has many grammatical reflexes. Thus, it looks like the distinction should be grammatically encoded. Various ways to encode this distinction into grammar have been suggested in the literature. A sample is provided in (41).

- (41) a. *Kratzer (1995)*: an i-predicate lacks a Davidsonian argument, contrary to s-predicates.
 b. *Chierchia (1995)*: an i-predicate bears a special feature that forces local agreement with a covert generic operator which mandatorily binds the Davidsonian argument of the i-predicate.
 c. *Diesing (1992)*: an i-predicate is selected by a special inflectional head that requires the subject of the i-predicate to be base generated directly in [Spec, IP], rather than in [Spec, VP] as it is the case for s-predicates.

Kratzer (1995) suggests that the s-/i-predicates distinction is encoded as a distinction in argumental structure, as in (41a). The ban in (2) and (3) against 'sometimes' or 'always' with i-predicates thus immediately follows from the fact that these adverbs have nothing to quantify over in the case of i-predicates. Chierchia (1995) suggests that the s-/i-predicates distinction is encoded as a distinction in syntactic features, as in (41b). Thus, the Davidsonian argument of an i-predicate is always bound by a covert operator and therefore inaccessible to overt adverbs, thus accounting for why 'sometimes' or 'always' are banned in the presence of i-predicates. Diesing (1992) suggests the further syntactic characterization (41c). She shows how to derive from this assumption the restriction on the readings of BPSs of i-predicates in (7) as well as the restriction on German word order in (9).

An alternative solution In chapter 3, I will explore a very different take on these facts. Let me take at face value the initial intuition that the i-predicates (1a) denote properties that are necessarily permanent, contrary to s-predicates. In other words, what is special about the i-predicate 'tall' is that it cannot happen that a given individual, say John, happens to be tall at some times in his life span but not at some others, i.e. the situation depicted in (42a) can never hold. Let me generalize

this situation by saying that an arbitrary unary *Predicate* is HOMOGENEOUS w.r.t. a *Restrictor* iff it cannot happen that some elements in the Restrictor satisfy the Predicate and some others don't, i.e. the situation depicted in (42b) can never hold. Thus, I am assuming that all that's special about the i-predicate 'tall' is that it is homogeneous w.r.t. John's life span.⁹



As noted at the end of section 1.1, a theory of i-predicates should account for why i-predicates cannot do the many things that s-predicates can do. If all that's special about i-predicates is that they are homogeneous, then there should be no properties peculiar to i-predicates; rather, all properties of i-predicates should extend to arbitrary homogeneous predicates. Chapter 3 argues that this prediction is borne out. I will consider various properties of i-predicates, such as those previewed in section 1.1, and I will show that each one of them formally corresponds to some of the cases considered in chapter 2, that contain a homogeneous predicate that is not an i-predicate. If indeed there are no specific properties of i-predicates, then specific assumption on i-predicates such as (41) cannot be on the right track, because they do not extend to arbitrary homogeneous predicates. In chapter 3, I will thus try to do without such specific assumptions and to derive instead the proper theory of i-predicates as a theorem of the more general theory of oddness independently developed in chapter 2. Let me preview the argument by discussing informally a couple of cases.

Adverbs Consider again the odd sentence (2b), repeated below in (43a), together with the odd sentence (44a), that has nothing to do with i-predicates.

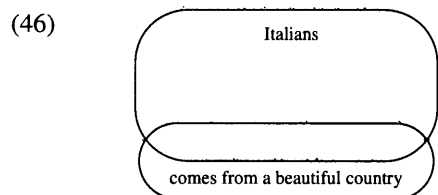
- (43) a. #John is sometimes tall.
b. John is tall.
- (44) a. #Some Italians come from a beautiful country.
b. Italians come from a beautiful country.

In chapter 2, I will suggest that the oddness of sentence (44a) can be accounted for by means of the informal piece of reasoning in (45), which is identical to the piece of reasoning in (22) used by Hawkins (1991) to account for the oddness of his examples (10).

- (45) a. On the one hand, sentence (44a) triggers the scalar implicature that the speaker cannot presume that all Italians come from a beautiful country.
b. On the other hand, common knowledge entails that all Italians come from the same country, hence (44a) cannot be true without all Italians coming from a beautiful country.
c. In conclusion, the oddness of sentence (44a) follows from the mismatch between the scalar implicature in (45a) and common knowledge in (45b).

The predicate 'to come from a beautiful country' is homogeneous w.r.t. to the restrictor 'Italians', in the sense that the situation depicted in (46), where some Italians come from a beautiful country while some others do not, cannot hold.

⁹ Actually, the scenario represented in (42a) mismatches with common knowledge for two different and unrelated reasons. The *first* reason is the one mentioned in the text, namely that (42a) depicts a scenario where there are some times in John's life span when he is tall and some others when he is not. The *second* reason is that (42a) depicts a scenario where there are times when John is tall but not alive. This second reason is completely irrelevant to my point. From this perspective, I should have actually made the set corresponding to "times when John is tall" be a (proper) subset of the set corresponding to "times when John is alive" in (42a). The reason why I have not drawn (42a) that way is to stress the analogy with other cases that have nothing to do with i-predicates. For instance, in the case of (46) I do not want the set corresponding to "comes from a beautiful country" to be a subset of the set corresponding to "Italians", because there might be other countries that are beautiful, besides Italy.



The two cases (43a) and (44a) are thus completely parallel: an existential quantifier binds the argument of a predicate that is homogeneous w.r.t. the corresponding restrictor. In chapter 3, I will take advantage of this parallelism and suggest that literally the same account (45) carries over to the oddness of sentence (43a), as in (47). Thus, the ban against existential quantifiers with *i*-predicates follows as a special case of the general ban against existential quantification over the argument of a homogeneous predicate.

- (47)
- a. On the one hand, sentence (43a) triggers the scalar implicature that the speaker cannot presume that John is always tall.
 - b. On the other hand, it follows from common knowledge that tallness is a property stable through time, hence (43a) cannot be true without John being always tall.
 - c. In conclusion, the oddness of (43a) follows from the mismatch between the scalar implicature (47a) and common knowledge (47b).

The parallelism between *i*-predicates and arbitrary homogeneous predicates extends to universal quantification. Consider the odd sentence (3b), repeated below in (48a), together with the odd sentence (49a), which has nothing to do with *i*-predicates.

- (48)
- a. #John is always tall.
 - b. John is tall.
- (49)
- a. #Every Italian comes from a beautiful country.
 - b. Italians come from a beautiful country.

In chapter 2, I'll suggest that the oddness of sentence (49a) can be accounted for by means of a piece of reasoning once more analogous to Hawkins' piece of reasoning (22): assume that the alternative (49b) contains a generic operator binding the variable introduced by the bare plural 'Italians'; following Fodor (1970), von Stechow (1997) and much subsequent literature, assume that this generic operator triggers the **HOMOGENEITY PRESUPPOSITION** that either all or no Italians come from a beautiful country; thus, an utterance of the odd sentence (49a) triggers the implicature that this homogeneity presupposition is false; but this implicature mismatches with the piece of common knowledge that all Italians come from the same country. In chapter 3, I will suggest that this same account straightforwardly carries over to the oddness of sentence (48a), since its alternative (48b) too contains a covert generic operator that quantifies over the time argument of the *i*-predicate 'tall' and carries an homogeneity presupposition that John is either always or never tall.

Bare plural subjects Consider the lack of the existential reading of the bare plural subject of sentence (7b), repeated below in (50a), together with the odd sentence (51a), which is a variant of an example discussed by Percus (2001), and has nothing to do with *i*-predicates.

- (50)
- a. Firemen are tall.
 - b. Some Firemen are tall.
- (51) *Context*: five different competitions were held separately, Monday through Friday; both John and Bill know that the same guy won all five of them:

	M	T	W	T	F
winner:	<i>x</i>	<i>x</i>	<i>x</i>	<i>x</i>	<i>x</i>

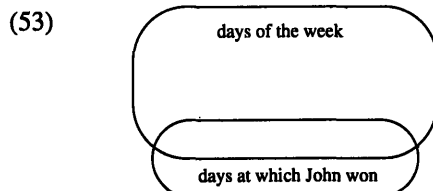
John wants to know more about this amazing guy, and thus asks Bill for more information. Bill replies as follows:

- a. #On every day, a/some fireman won.
 b. A/Some fireman won on every day.

Contrary to the case of the fine sentence (51b), the existential quantifier over firemen in the odd sentence (51a) has narrow scope with respect to the universal quantifier over days, because of overt fronting of the latter. In chapter 2, I will derive the oddness of sentence (51a) from this scope configuration by means of the informal piece of reasoning in (52), which is identical to the piece of reasoning in (22) used by Hawkins (1991) to account for the oddness of his examples (10). For the time being, let me not worry on how exactly the scalar implicature in (52a) is derived.

- (52) a. On the one hand, an utterance of (51a) triggers the implicature that the stronger alternative (51b), with the opposite scope configuration, is false;
 b. On the other hand, common knowledge entails that the same guy always won, hence (51a) cannot be true without (51b) being true too.
 c. In conclusion, the oddness of (51a) follows from the mismatch between the scalar implicature in (52a) and the common knowledge in (52b).

The predicate 'win' is homogeneous in the context considered in (51), in the sense that the situation depicted in (53), where a given individual, say John, won on some days of the week but not on some others, cannot hold. Furthermore, it is crucial that the indefinite 'a/some fireman' has scope below the universal quantifier 'on every day' in the odd sentence (51a), since sentence (51b), where the scope is reversed, is acceptable.



The two cases (50) and (51) are thus completely parallel. In fact, the predicate 'tall' is homogeneous with respect to the time argument. Furthermore, the alleged existential bare plural subject of sentence (50a) could only have the narrowest scope, because of a general tendency of existential bare plurals to always take the narrowest possible scope, exemplified by well known contrasts such as the one in (54), from Carlson (1977).

- (54) a. Some dogs hung around my house all last week. \exists -dogs > all last week.
 b. Dogs hung around my house all last week. $*\exists$ -dogs > all last week.

In chapter 3, I will take advantage of this parallelism and suggest that literally the same account (52) carries over to the lack of the existential reading of the bare plural subject of the i-predicate of sentence (50a). I give a rough sketch of the reasoning in (55). Here, I am assuming that, if a sentence is ambiguous and one of its two readings is odd, the sentence is fine but the odd reading gets hidden.

- (55) a. The existential bare plural subject of sentence (50a) has narrow scope with respect to the operator that binds the time argument of the predicate, because that is always the case with existential bare plurals; thus, sentence (50a) with the bare plural subject construed existentially triggers the scalar implicature that the corresponding sentence (50b) with wide scope existential quantifier over firemen is false.
 b. But common knowledge entails that, if at every time there is a fireman who is tall at that time, then there has got to be a fireman who is tall at every time.
 c. The unavailability of the existential reading of the bare plural subject of (50a) thus follows from the mismatch between the scalar implicature (55a) and the common knowledge (55b).

In chapter 3, I will show that this account extends to various other facts concerning bare plural subjects of i-predicates, such as Fox's contrast (8) and the restriction on German word order (9).

1.4 Implications

The following two claims (56) have been defended in the recent literature. Chierchia (2006b) and Fox (2007a) among others defend claim (56a): they suggest that scalar implicatures arise because of a syntactically realized exhaustivity operator. Chierchia (2004) and Krifka (1995) among others illustrate (56b): they suggest an account for NPIs and their intervention effect based on scalar implicatures. Part I of my dissertation contributes to both claims (56).

- (56) a. Scalar implicatures must be derived within grammar, rather than in the extra-grammatical domain of pragmatics.
- b. A grammatical theory of scalar implicatures can be used to provide a semantic account for various grammatical facts.

In chapter 2, I will contribute to claim (56a). I will discuss various cases of odd sentences and I will argue that their oddness follows from the fact that they trigger scalar implicatures that mismatch with common knowledge, as initially proposed by Hawkins' account (22) for the oddness of sentences (10). Building on a remark by Heim (1991), I have noted that, in order for this account to work, it is crucial that strengthening be blind to common knowledge. And that this hypothesis makes no sense within the pragmatic theory of scalar implicatures. I will try to make sense of this hypothesis within a framework where scalar implicatures arise through a syntactically realized operator. The debate between the pragmatic and the grammatical approaches to scalar implicatures is currently alive in the literature: see Sauerland (2004c), Russell (2004), Horn (2005), Geurts (2009) among others, for a defense of the original pragmatic approach; see Fox (2007a), Chierchia (2007) and Chierchia et al. (2007) for a defense and various implementations of the grammatical approach. To the extent that my defense of Hawkins' proposal will be successful, it will contribute to this debate, providing evidence in favor of the grammatical framework and against the pragmatic one. In chapter 3, I will contribute one more example of (56b). As sketched in (41), various facts pertaining to *i*-predicates have been so far accounted by dedicated grammatical assumptions. I will suggest that all that's special about *i*-predicates is that it follows from common knowledge that they are homogeneous w.r.t. time. To make my point, I will try to argue that various properties of *i*-predicates extend to arbitrary homogeneous predicates. And I will derive various properties of *i*-predicates from the version of the theory of scalar implicatures independently defended in chapter 2. I will thus provide one more example of an account based on scalar implicatures for facts that had been so far accounted for through core grammar.

Chapter 2

Oddness by mismatching scalar implicatures

Sentence (57) sounds odd because it *somehow* mismatches with the piece of common knowledge that all Italians come from the same country. How does this mismatch happen? As reviewed in section 1.2, this question is not trivial, and it has been indeed the topic of an intense debate in the very recent literature. In this chapter, I will address this question.

(57) #Some Italians come from a beautiful country.

To illustrate my take on this question, let me turn to the case of the odd sentence (58). I think that there can hardly be any debate on the source of the oddness of this sentence: it sounds odd because it *says*¹ that some but not all Italians come from a beautiful country, which is false in every world compatible with the piece of common knowledge that all Italians come from the same country. This has got to be the reason for the oddness of sentence (58). No debate in this case!

(58) #Only some_F Italians come from a beautiful country.

A recent theory of scalar implicatures says that scalar implicatures are not a post-grammatical, pragmatic phenomenon. Rather, scalar implicatures are derived in the syntax by appending to LFs a phonologically covert variant of ‘only’, notated EXH and called the EXHAUSTIVITY OPERATOR. Thus, the pair of (57) and (58) can be made more explicit as in (59). Faced with the parallelism in (59), it is tempting to say that the initial sentence (57) sounds odd for *exactly* the same obvious reason why sentence (58) sounds odd. Namely, because its LF (59a) *says* that some but not all Italians come from a beautiful country. This is of course just a restatement of Hawkins’ account (22).

- (59) a. #EXH some Italians come from a beautiful country.
b. #Only some Italians come from a beautiful country.

The goal of this chapter is to investigate how far does this intuitive idea get. Section 2.1 starts with simple cases and formalizes the idea: it reviews from the literature the idea that scalar implicatures are derived in the grammar by means of a covert exhaustivity operator and addresses the Blindness issue (25)/(27) and the Mandatoriness issue (26)/(28) on this theoretical background. Section 2.2 turns to cases where mismatch with common knowledge arises not from a single implicature but rather from the coordinated action of a whole bunch of implicatures. Section 2.3 discusses cases where the scalar item that triggers the mismatching implicature is embedded underneath some other operator, paying particular attention to the case where the embedding operator is downward entailing. Section 2.4 explores the hypothesis that anti-presuppositions should be subsumed under the general case of scalar implicatures. Finally, section 2.5 collects various miscellaneous issues. None of the examples considered in this chapter have anything to do with *i*-predicates. In chapter 3, I will turn to *i*-predicates and argue that their properties can be derived as a theorem of the more general

¹Ignoring the issue of the division of labor between presupposition and assertion in the proper semantics of ‘only’.

theory of oddness developed in this chapter, since each one of the properties of *i*-predicates considered in the next chapter will turn out to formally correspond to one of the examples studied in this chapter.

2.1 The basic case

The two sentences in each pair (60)-(64) are contextually equivalent. Yet, the (a) sentences with ‘some’ sound odd while the contextually equivalent (b) sentences with the definite sound fine.² Example (61) is adapted from Percus (2001); example (62) is due to a (p.c.) by E. Chemla reported in Schlenker (2006); example (63) is due to Spector (2007a).

- (60) Mary is conducting a survey on names and last names of Italian children. She knows that all children inherit the last name of their father; hence, all children of a given couple share the same last name. This week, she has interviewed the children of five couples, A through E, in order to record their names and last names...
- a. #Some children of couple C have a long last name.
 - b. The children of couple C have a long last name.
- (61) Alluding to the Marx Brothers:
- a. #The mother of one of them was named Minnie.
 - b. Their mother was named Minnie.
- (62) In a department where every professor assigns the same grade to all of his students:
- a. #This year, prof. Smith gave an A to some of his students.
 - b. This year, prof. Smith gave an A to his students.
- (63) Speaking of the atoms of a molecule which cannot be separated:
- a. #Some atoms went right.
 - b. The atoms went right.
- (64) a. #Some Italians come from a beautiful country.
b. Italians come from a beautiful country.

The two sentences in (65), from Fox (2004), are contextually equivalent, since common knowledge ensures that (65a) with the small numeral is equivalent to (65b) with the large numeral. Yet, (65a) with the small numeral sounds odd while (65b) with the large numeral sounds fine.

- (65) John has an odd number of children...
- a. #He has two children.
 - b. He has three children.

The two sentences in the two pairs (66) and (67) are contextually equivalent, since common knowledge ensures that the (a) sentences with disjunction or with a single disjunct are equivalent to the corresponding (b) sentences with conjunction. Yet, the (a) sentences with disjunction or with a single disjunct sound odd while the equivalent (b) sentences with conjunction sound fine. Example (66) is due to Hurford (1974).

- (66) John and Mary travelled from Vienna to Paris together...
- a. #John or Mary travelled by train.
 - b. John and Mary traveled by train.
- (67) John and Mary travelled from Vienna to Paris together...
- a. #John traveled by train.

²For the case of (64), I assume that sentence (64b) contains a generic operator which is akin to the definite article; see Ferreira (2005) for a discussion of parallelisms between the definite article and the generic operator. The pattern is more uniform in Italian, where sentence (64b) would have a definite.

- b. John and Mary traveled by train.

In this section, I formalize the idea that the (a) sentences just listed feel odd because they contain a covert ‘only’. Subsection 2.1.1 reviews from the literature the hypothesis of a covert ‘only’. Subsection 2.1.2 addresses the Blindness issue (25)/(27). Subsection 2.1.3 addresses the Mandatoriness issue (26)/(28).

2.1.1 First step: covert ‘only’

According to the neo-Gricean framework, grammar pairs each sentence with a PLAIN MEANING and scalar implicatures are then computed on the basis of this meaning by an extra-grammatical procedure rooted in general-purpose principles of rationality. Chierchia (2004), building on Landman (1998), Levinson (2000) and Cohen (1971), questions this approach by arguing that it fails to account for a pattern of implicatures in embedded contexts. He then goes on to sketch a very different framework, whereby scalar implicatures are derived through a purely grammatical algorithm, to be conceived of as a grammaticalization of the standard Gricean reasoning. Thus, grammar pairs each sentence with both a plain meaning and a STRENGTHENED MEANING, namely the plain meaning enriched with its scalar implicatures. Fox (2007a) notes that the strengthened meaning of sentences (68a), (69a) and (70a) can be paraphrased as the plain meaning of the corresponding sentences (68b), (69b) and (70b), with overt ‘only’.

- (68) a. John bought three houses.
 b. John only bought three_F houses.
- (69) a. John did some of the homework.
 b. John only did some_F of the homework.
- (70) a. John talked to Mary or Sue.
 b. John only talked to Mary or_F Sue.

The situation in (68)-(70) is similar to the one in (71). English has an overt distributivity operator ‘each’, that makes it possible in (71b) to apply the distributive predicate denoted by ‘tall’ to the plural individual denoted by the definite subject ‘the kids’. Sentence (71a), without the overt operator ‘each’, sounds completely equivalent to the corresponding sentence (71b). A standard strategy to account for this parallelism is to assume that English has a phonologically covert variant of ‘each’, called the DISTRIBUTIVITY OPERATOR and notated DIST. The LF of sentence (71a) contains this covert operator and is thus identical to the LF of sentence (71b).

- (71) a. The kids are tall.
 b. The kids each are tall.

Fox (2007a) suggests that the parallelism in (68)-(70) should be handled in quite the same way. He assumes that Natural Language has a phonologically covert variant of ‘only’, called the EXHAUSTIVITY OPERATOR and notated EXH. And that the strengthened meaning of the sentences (68a)-(70a) is the plain meaning of the corresponding LFs adorned with this special covert propositional operator EXH, as stated in (72). Various arguments have been provided in the literature in favor of this assumption (72). One argument is that (72) allows for the exhaustivity operator to appear in embedded contexts, thus straightforwardly accounting for various patterns of embedded implicatures; see Fox (2007a) and Chierchia et al. (2007). Another argument is that (72) allows for the the exhaustivity operator to be iterated, as noted in Spector (2006), Spector (2007a) and Fox (2007a).

- (72) strengthened meaning of sentence $\varphi = \llbracket [\text{EXH } \varphi] \rrbracket$.

Many approaches to the semantics of ‘only’ or, equivalently, of the exhaustivity operator EXH share the structure in (73):³ EXH(φ) asserts φ and furthermore negates a bunch of alternatives ψ , namely all the alternatives ψ in the set $\mathcal{E}xcl(\varphi)$ of ALTERNATIVES EXCLUDABLE W.R.T. φ . Each conjunct

³Here and throughout the paper, I sloppily use the same symbol φ for both an LF and its truth conditions.

$\neg\psi$ in (73) is called a SCALAR IMPLICATURE of the prejacent φ . The set $\mathcal{Excl}(\varphi)$ of alternatives excludable w.r.t. φ is a subset of the set $\mathcal{Alt}(\varphi)$ of SCALAR ALTERNATIVES of φ . Thus, these various definitions of the exhaustivity operator differ in how they define the two sets $\mathcal{Alt}(\varphi)$ and $\mathcal{Excl}(\varphi)$.

$$(73) \quad \text{EXH}(\varphi) = \varphi \wedge \bigwedge_{\psi \in \mathcal{Excl}(\varphi)} \neg\psi.$$

The standard definition of the set $\mathcal{Alt}(\varphi)$ of the scalar alternatives of a sentence φ is the one in (74), whereby scalar alternatives of φ are obtained by arbitrarily replacing scalar items in φ with their Horn-mates. I will present my assumptions on Horn-scales one at the time, as I need them.

(74) The set $\mathcal{Alt}(\varphi)$ contains all and only those ψ 's that can be obtained from φ by replacing one or more scalar items in φ with their Horn-mates.

The classical definition of the set $\mathcal{Excl}(\varphi)$ of alternatives excludable w.r.t. a sentence φ is (75), whereby $\mathcal{Excl}(\varphi)$ is the set of all the scalar alternatives ψ 's which asymmetrically entail φ .

$$(75) \quad \mathcal{Excl}(\varphi) = \{\psi \in \mathcal{Alt}(\varphi) \mid \psi \rightarrow \varphi, \varphi \not\rightarrow \psi\}.$$

In section 2.2, I will pause to consider refinements of the current definition (74) of the set of scalar alternatives and of the current definition (75) of the set of excludable alternatives, suggested in Fox (2007a).

2.1.2 Second step: blindness

In this subsection, I discuss the Blindness issue (25)/(27) for Hawkins' original proposal (22). As noted above, the Gricean or neo-Gricean framework is incompatible with the hypothesis that scalar implicatures are computed blind to common knowledge: if the computation of implicatures is driven by general-purpose principles of rationality, then it cannot be encapsulated in such a way as to ensure blindness to common knowledge. In this section, I try to suggest that the opposite holds true for the grammatical theory of scalar implicatures, namely that this framework is incompatible with the hypothesis that scalar implicature are computed taking common knowledge into account. My discussion follows closely Fox and Hackl (2006, Sect. 5).

The issue The definition of the exhaustivity operator EXH introduced in the preceding subsection makes use of the notion of entailment. To complete the definition of the exhaustivity operator, I thus need to spell out the relevant notion of entailment. Let \mathcal{W} be the set of all possible worlds. Let \mathcal{W}_{ck} be the subset of those possible worlds where COMMON KNOWLEDGE holds. Common knowledge can be of two types. It can be true, realistic, well entrenched common knowledge, such as the piece of knowledge that people have two eyes of the same color or that all Italians come from the same country. Or it can be an arbitrary piece of knowledge turned into shared knowledge by virtue of some preceding text. Throughout the paper, I will not distinguish between oddness that arises because of mismatch against one or the other of these two types of common knowledge, but I will always try to illustrate each pattern with examples of both kinds.⁴ Once the two sets of possible worlds \mathcal{W} and \mathcal{W}_{ck} are in place, the notion of entailment can be spelled out in two different ways: as LOGICAL entailment or as entailment GIVEN COMMON KNOWLEDGE \mathcal{W}_{ck} . The definitions are provided in (76). Of course, both notions of entailment amount to set-inclusion. The difference between the two notions is that set-inclusion is only checked within \mathcal{W}_{ck} in the case of (76b). I denote logic entailment by " $\rightarrow_{\mathcal{W}}$ " and entailment given common knowledge by " $\rightarrow_{\mathcal{W}_{\text{ck}}}$ ".

(76) For any two propositions φ, ψ :

a. ψ LOGICALLY entails φ iff $\psi \subseteq \varphi$.

⁴In many cases, it looks like mismatch against a realistic, well entrenched piece of common knowledge yields a "stronger" oddness effect than mismatch against some piece of knowledge introduced by a preceding text. My proposal does not attempt at modeling this gradience. I implicitly assume that the intensity of the oddness effect is a function of the entrenchment of the corresponding piece of common knowledge. The degree of the latter is not a matter of linguistics. Thus, the degree of the oddness effect is not a matter of linguistics either.

- b. ψ entails φ GIVEN COMMON KNOWLEDGE \mathcal{W}_{ck} iff $\psi \cap \mathcal{W}_{ck} \subseteq \varphi$.

I will come back to the issue of the proper definition of logic entailment in subsection ???. Which one of these two notions of entailment is the one relevant for the definition of the exhaustivity operator EXH? This subsection addresses this question.

The idea Assume that scalar implicatures are indeed computed by a grammatical module. This would have to be a grammatical module sensitive to some notion of entailment. We can thus address the question just raised by looking at which notion of entailment is used by other entailment-sensitive grammatical modules. It turns out that there is some evidence that such modules are sensitive to logic entailment rather than to contextual entailment; let me quickly review two such examples, provided in Fox (2000, §2.5) and Gajewski (2003). Fox (2000) argues that covert movement is subject to the constraint SCOPE ECONOMY, which roughly says that a quantifier can undergo covert movement in English only if the pre-movement LF and the post-movement LF are not equivalent.⁵ This statement of course raises the question of whether the relevant notion of equivalence is logic equivalence or equivalence relative to common knowledge. Fox provides evidence that Scope Economy must be stated in terms of logical equivalence, not in terms of equivalence given common knowledge. Thus, the module which checks Scope Economy is an example of an entailment-sensitive grammatical module which is blind to common knowledge. Gajewski (2003) provides a second example of an entailment-sensitive grammatical module blind to common knowledge. He reviews some recent literature where a CONTRADICTION FILTER is used to account for the deviance of a sentence on the basis of its contradictoriness; see for instance Chierchia (1988, Sect. II.3.3), Barwise and Cooper (1991) and von Stechow (1993); and see Ladusaw (1986) for discussion. For instance, von Stechow (1993) derives the deviance of the ‘but’-phrase in (77a) from the fact that the sentence is necessarily contradictory, given a proper semantics for ‘but’-phrases. Yet, Gajewski notes that contradictoriness in general does not result in deviance, as in the case of the embedded clause in (77b).⁶ What is the relevant difference between the two cases (77a) and (77b)? Gajewski suggests that the Contradiction Filter is checked at a level of representation where only logical operators are retained, while non-logical lexical entries are ignored, together with the common knowledge that they carry along. At this special level of representation, only sentence (77a) yields a contradiction, while the embedded clause of (77b) does not. Thus, the module that checks the Contradiction Filter is another example of an entailment-sensitive grammatical module which is blind to common knowledge.

- (77) a. #Some students but Bill passed the exam.
 b. Your assumptions entail that [Socrates is mortal and immortal].

I thus conclude that there is some independent evidence that entailment-sensitive grammatical algorithms are blind to common knowledge. Assume now that scalar implicatures are not pragmatic inferences but rather are computed by a dedicated entailment-sensitive grammatical algorithm, as suggested by the Grammatical framework for scalar implicatures. Then it makes sense to assume that the latter algorithm should be blind to common knowledge too, namely that the notion of entailment relevant for the computation of scalar implicatures is that of logic entailment (76a) rather than that of contextual entailment (76b).

The proposal So far, I have considered the blindness issue from the *general* perspective of the grammatical framework to scalar implicatures. Now I want to concentrate on the specific instance of the grammatical framework reviewed in subsection 2.1.1. According to this specific proposal, scalar implicatures are derived by appending to the LF a covert exhaustivity operator EXH. This operator is construed as a covert variant of overt ‘only’. It thus makes sense to assume that the two operators differ from each other only minimally. For the sake of the argument, let me take the analogy perhaps a step too far, and assume that (73) is also the proper semantics of overt ‘only’, as stated in (78). It thus makes some sense to try to address the question of whether the exhaustivity operator is blind to

⁵See (373) below for a more precise formulation.

⁶Which is a variant of Gajewski’s examples considered in Fox and Hackl (2006).

common knowledge by addressing the corresponding question for overt ‘only’, then extending the solution to the covert exhaustivity operator by analogy.

$$(78) \quad \llbracket \text{only} \rrbracket(\varphi) = \varphi \wedge \bigwedge_{\psi \in \mathcal{E}xcl(\varphi)} \neg\psi.$$

Let’s consider again the odd sentence (57b) with overt ‘only’, repeated below in (79). Let φ be the prejacent of ‘only’, as in (79a); let ψ be the corresponding alternative with ‘some’ replaced by ‘all’, as in (79b). This sentence mismatches with the piece of common knowledge that all Italians come from the same country. The most straightforward account for this mismatch is as follows: by (78), the plain meaning of sentence (79) entails the negation of the alternative ψ . Yet, if the computation of the set $\mathcal{E}xcl(\varphi)$ of excludable alternatives in (78) could take common knowledge into account, then the prejacent φ with ‘some’ and the alternative ψ with ‘every’ would be equivalent and thus the latter alternative ψ would not be excludable. I thus conclude that the proper computation of the set of excludable alternatives for overt ‘only’ must be blind to common knowledge.

- (79) #Only some_F Italians come from a beautiful country.
- a. φ = Some Italians come from a beautiful country.
 - b. ψ = All Italian come from a beautiful country.

Yet, this conclusion is threatened by the following alternative account. Assume that the computation of the set of excludable alternatives for overt ‘only’ is not at all blind to common knowledge. Thus, the alternative ψ in (79b) does *not* belong to the set of excludable alternatives of the prejacent φ in (79a). The oddness of sentence (79) is thus not due to any mismatch with common knowledge. Rather, the oddness of sentence (79) can be explained as follows. Since ψ in (79b) is not excludable, then the set of excludable alternatives is empty in the case of (79). Overt ‘only’ is therefore vacuous. And sentence (79) is ruled out by the same general constraint that bans the vacuous occurrence of ‘only’ in sentence (80).

- (80) #Only every_F boy arrived.

Yet, I think that this alternative explanation can be counterattacked by considering cases with multiple alternatives, such as (81). Suppose that the set of excludable alternatives of overt ‘only’ is computed taking into account the common knowledge that John has an odd number of children. In this case, the alternative ψ that John has (at least)⁷ three children is not excludable, since it is equivalent to the prejacent φ that John has (at least) two children in the context considered. Yet, the occurrence of ‘only’ in this sentence (81) is in no way vacuous, because it can still negate the alternative ψ' that John has (at least) four children. Thus, the hypothesis that the semantics of ‘only’ is sensitive to common knowledge leads to the incorrect prediction that the sentence ‘John has only two’ should be fine in the context considered, and furthermore should mean that John has exactly three children!

- (81) John has an odd number of children. . .
 . . . #He has only two_F.
- a. φ = John has at least two.
 - b. ψ = John has at least three.
 - c. ψ' = John has at least four.

Based on these considerations, I conclude that the set of excludable alternatives for overt ‘only’ is computed using logic entailment (76a) and is therefore blind to common knowledge. By virtue of the analogy between overt ‘only’ and the covert exhaustivity operator EXH, I thus suggest (82), namely

⁷I take the fact that ‘only’ can be construed with a numeral as evidence that numerals do have a weak semantics, as shown by the contrast in (i)

- (i) a. John has only two_F children.
 b. #John has only exactly two_F children.

that the computation of the strengthened meaning $\text{EXH}(\varphi)$ of a sentence φ is blind to common knowledge. I'll thus dub (82) the BLINDNESS HYPOTHESIS (henceforth: BH).

- (82) BLINDNESS HYPOTHESIS. The notion of entailment relevant for the definition of the exhaustivity operator EXH is that of logical entailment rather than that of entailment given common knowledge \mathcal{W}_{ck} .

I started out with the intuition (25) that the BH (82) cannot be true, because the corresponding theory of implicatures would sound somewhat paradoxical. And I have concluded that the BH (82) cannot be false, because the corresponding theory of implicatures would paradoxically predict the sentence 'John has (only) two children' to be able to mean in some circumstances that he has exactly three!

2.1.3 Third step: mandatoriness

In this subsection, I discuss the Mandatoriness issue (26)/(28) for Hawkins' original proposal (22). More precisely, I spell out a rather straightforward account of the context-dependence of scalar implicatures within the version of the grammatical theory of scalar implicatures reviewed in subsection 2.1.1. And I argue that this account for context-dependence is compatible with the surprising mandatoriness of those special scalar implicatures that happen to mismatch with common knowledge.

The issue Let's say that a scalar implicature is *mandatory* if it is generated by default without any triggering effect of context. Granted that implicatures are indeed computed blind to common knowledge according to the BH (82), sentence (57a), repeated in (83a), *may* trigger the scalar implicature that the logically stronger alternative (83b) is false, which mismatches with the piece of common knowledge that all Italians come from the same country.

- (83) a. #*Some* Italians come from a beautiful country.
b. *All* Italians come from a beautiful country.

But of course, this is not enough: in order to derive the oddness of sentence (83a), I need to assume that this mismatching implicature *must* be computed, namely that this mismatching implicature is mandatory. Yet, this hypothesis is at odds with the very well known fact that implicatures in general are *not* mandatory. To illustrate, consider for instance sentence (84c): as an answer to the question (84a), it triggers the scalar implicature that B doesn't know whether John is available before dinner; but as an answer to question (84b), it doesn't quite seem to implicate that, but rather that no one else besides John is available after dinner.

- (84) a. A: When is John usually available?
b. A': Who is usually available after dinner?
c. B: John is usually available after dinner.

In conclusion, I need a way to reconcile the alleged mandatoriness of mismatching implicatures with the fact that standard non-mismatching implicatures are not mandatory at all.⁸

The idea Here is a possible way to go.⁹ Scalar alternatives can be relevant or not in a given context. Of course, if a scalar alternative ψ is not relevant, then it triggers no implicature $\neg\psi$. In the context

⁸A completely analogous problem arises for Chierchia's (2004) account for the intervention effect with NPI's in terms of scalar implicatures. Chierchia (2006b) offer a syntactic solution to the problem. As far as I can see, his proposal does not extend to my case.

⁹An alternative way to go, suggested to me by Benjamin Spector (p.c.), would be to exploit the distinction between primary and secondary implicatures, introduced by Sauerland (2004c). PRIMARY scalar implicatures are ignorance inferences of the form "it is not the case that the speaker believes that..." (in brief: $\neg\mathcal{B}_{\text{speaker}}\psi$). SECONDARY scalar implicatures have the opposite scope between negation and the belief operator, namely have the form "the speaker believes that it is not the case that..." (in brief: $\mathcal{B}_{\text{speaker}}\neg\psi$). Sauerland (2004b) notes the contrast in (i): sentence (ib) cancels only the secondary implicature of sentence (ia), namely that the speaker knows that not all Beethoven's symphonies were played; sentence (ic) cancels also the primary implicature, namely that it is not the case that the speaker believes that all symphonies were played; the contrast in acceptability between (ib) and (ic) suggests that secondary implicatures can be canceled by the simple assertion of the opposite while primary implicatures cannot. This observation seems to suggest that primary implicatures are harder to

of question (84a), the alternative ‘Mary is available after dinner’ is not relevant, hence it triggers no implicature; in the context of question (84b), the alternative ‘John is available before dinner’ is not relevant, hence it triggers no implicature. Furthermore, let me assume that also the converse holds: if a scalar alternative ψ is relevant, then its corresponding implicature $\neg\psi$ is mandatory. From this perspective, implicature cancellation, as in (85), really amounts to switching from a context where the alternative ‘he read all’ is not relevant to a different one, where it is; see van Kuppevelt (1996) for relevant discussion.

- (85) a. Q: Who read some of the books?
 b. A: John read some, in fact he read all.

Assume furthermore that, precisely because of its contextual nature, relevance cannot distinguish between two propositions which are contextually equivalent: they are either both relevant or else neither of them is. What’s special about the mismatching alternative (83b) is that it *must* be relevant in the context considered, since it is contextually equivalent to the target utterance (83a), which is relevant precisely because of the fact that it has been uttered. Since the mismatching alternative (83b) is necessarily relevant in the context considered, the corresponding implicature is mandatory and oddness is thus accounted for. In conclusion, scalar implicatures are of course in general not mandatory, because they might correspond to irrelevant alternatives; but mismatching implicatures are indeed mandatory, because they correspond to alternatives which are always relevant, because contextually equivalent to the assertion.

The proposal: first part For the sake of explicitness, let me offer a concrete implementation of the idea I have just sketched. Within the framework adopted here, scalar implicatures of a sentence φ are derived by appending to the LF of the sentence a covert variant EXH of overt ‘only’. The assumption of a covert ‘only’ immediately raises the following question: how does the hearer recover whether the LF produced by the speaker contains that covert operator or not? For the time being, let’s restrict ourselves to the case of matrix sentences; I will come back to the case of embedded scalar items in section 2.3. With this restriction, the simplest possible answer to our question would be (86): it says that there is no recoverability problem for the exhaustivity operator, since it is mandatorily present.¹⁰

cancel than secondary implicatures, and thus feel more robust.

- (i) a. They played many of Beethoven’s symphonies, ...
 b. ... and possibly all.
 c. ... ?and definitely all.

One possible strategy to cope with the mandatoriness issue is thus as follows: secondary implicatures are context dependent, but primary implicatures are not, namely they are mandatory; and the mismatching implicatures needed for Hawkins’ proposal are primary implicatures. I have not substantially explored this alternative account because I fear that it might be incompatible with the proposal I will make in section 2.2.

¹⁰Here is one immediate problem for this assumption (86). Consider the contrast in (i): sentence (ia) turns deviant if we add an overt ‘only’ associated with the universal determiner ‘every’ as in (ib). The reason has intuitively to lie with the fact that ‘only’ in (ib) does not contribute to meaning, since it is associated with the logically strongest alternative in the focus set.

- (i) a. Every boy arrived.
 b. ?Only every_F boy arrived.

The contrast in (i) might be problematic for my assumption (86), as follows. By the latter assumption, the fine sentence (ia) contains a mandatory exhaustivity operator, so that the pair in (i) should really be construed as the pair in (ii). If the semantics of the covert exhaustivity operator is indeed akin to that of overt ‘only’, then the contrast in (i) is puzzling.

- (ii) a. EXH every boy arrived.
 b. ?Only every boy arrived.

There are two possible ways out. One way is to exploit the obvious fact that ‘only’ is overt while EXH is covert. If we had some kind of BREVITY PRINCIPLE, then we might try to construe it in such a way that it only applies to overt items and not to covert items too, thus ruling out (iib) but not (iia). As discussed in subsection 2.5.1, I find such a principle very hard to formulate. An alternative way to deal with the contrast in (ii) is to exploit possible differences between the semantics of overt ‘only’ and that of covert EXH and thus to attribute the deviance of (iib) to some peculiar aspect of the semantics of ‘only’. For instance, suppose that the crucial difference between the two operators has to do with the proper division of labor between assertion and presupposition. Following a long tradition, let me assume that ‘only’ presupposes its prejacent, thus

(86) The exhaustivity operator EXH is mandatorily adjoined to every matrix clause.

Once assumption (86) is in place, special care is needed in order to account for the context sensitivity of scalar implicatures, as illustrated above in (84). Here is a way to go. According to Fox (2007a), the exhaustivity operator EXH has the shape in (73), repeated in (87a). According to this definition, the exhaustivity operator EXH takes a proposition φ and returns that same proposition conjoined with the negation of *all* the alternatives ψ 's which are in the set $\mathcal{E}xcl(\varphi)$ of alternatives excludable given φ . Following Fox and Katzir (2009), let me replace this definition (87a) with the slight modification in (87b). According to this modification, the exhaustivity operator depends on a contextually-supplied “question under discussion” \mathcal{R} of type $\langle\langle s, t \rangle, t\rangle$: \mathcal{R} is a property of propositions which holds of a proposition ψ iff ψ is relevant in the given context. According to the definition (87b), the exhaustivity operator $EXH_{\mathcal{R}}$ takes a proposition φ and returns that same proposition enriched with the claim that each excludable alternative ψ is either false or irrelevant. Of course, (87b) is equivalent to (87b'). Thus, I can equivalently restate as follows: the exhaustivity operator $EXH_{\mathcal{R}}$ takes a proposition φ and returns that same proposition conjoined with the negation of all the alternatives ψ 's which are in the set $\mathcal{E}xcl(\varphi)$ of alternatives excludable given φ and are furthermore relevant according to \mathcal{R} . In this section, I use the formulation (87b); in the next section 2.3, I will start from the formulation (87b') and present a more radical version of this proposal.

$$(87) \quad \begin{array}{l} \text{a. } EXH(\varphi) = \varphi \wedge \bigwedge_{\psi \in \mathcal{E}xcl(\varphi)} \neg\psi. \\ \text{b. } EXH_{\mathcal{R}}(\varphi) = \varphi \wedge \bigwedge_{\psi \in \mathcal{E}xcl(\varphi)} (\neg\psi \vee \neg\mathcal{R}(\psi)). \\ \text{b'. } EXH_{\mathcal{R}}(\varphi) = \varphi \wedge \bigwedge_{\psi \in \mathcal{E}xcl(\varphi) \cap \mathcal{R}} \neg\psi. \end{array}$$

Let me now illustrate how this proposal works. Consider sentence (84c), repeated as φ in (88a). Assume that it comes with the two scalar alternatives ψ_1 and ψ_2 in (88b) and (88c). According to (86), the exhaustivity operator is mandatory. According to (87b), the exhaustivity operator yields the strengthened meaning in (88), dependent on the contextual variable \mathcal{R} . Note that the shape of the strengthened meaning (88) is the same no matter whether the sentence φ is an answer to question (84a) or to question (84b). Yet, the variable \mathcal{R} in the strengthened meaning (88) will be assigned different values in these two contexts (84a) and (84b). In the context of question (84a), the alternative ψ_1 is relevant but the alternative ψ_2 is not; hence, the strengthened meaning in (88) boils down to $EXH_{\mathcal{R}}(\varphi) = \varphi \wedge \neg\psi_1$, namely to the proposition that John is available only after dinner. In the context of question (84b), the alternative ψ_2 is relevant but the alternative ψ_1 is not; hence, the strengthened meaning in (88) boils down to $EXH_{\mathcal{R}}(\varphi) = \varphi \wedge \neg\psi_2$, namely to the proposition that only John is available after dinner.

replacing the initial assumption (78) with the revised assumption (iia); see for instance Simons (2000) and Ippolito (2006) for a discussion of assumption (iia). Let me assume furthermore that EXH is a non-presuppositional variant of ‘only’, as in (iib).

$$(iii) \quad \begin{array}{l} \text{a. } [\text{only}] = \lambda\varphi : \varphi \wedge \bigwedge_{\psi \in \mathcal{E}xcl(\varphi)} \neg\psi. \\ \text{b. } EXH = \lambda\varphi . \varphi \wedge \bigwedge_{\psi \in \mathcal{E}xcl(\varphi)} \neg\psi. \end{array}$$

No matter whether sentence (iia) contains or not the exhaustivity operator, its presupposition and assertion will boil down to (iva). The case of (iib) is different, since its presupposition and assertion are now (ivb).

- | | | | |
|------|----|---------------------------------|--------------------|
| (iv) | a. | <i>presupposition of (iia):</i> | tautology. |
| | | <i>assertion of (iia):</i> | every boy arrived. |
| | b. | <i>presupposition of (iib):</i> | every boy arrived. |
| | | <i>assertion of (iib):</i> | tautology. |

I could now try to account for the contrast in (ii) as follows. Sentence (iib) with overt ‘only’ is deviant because “unbalanced”: as noted in (ivb), it asserts nothing and only carries a presupposition. Sentence (iia) with covert EXH is fine because “well balanced”: as noted in (iva), all the content is presented as an assertion and not as a presupposition.

- (88) $\text{EXH}_{\mathcal{R}}(\varphi) = \varphi \wedge (\neg\psi_1 \vee \neg\mathcal{R}(\psi_1)) \wedge (\neg\psi_2 \vee \neg\mathcal{R}(\psi_2))$.
- $\varphi =$ John is available after dinner.
 - $\psi_1 =$ John is available before and after dinner.
 - $\psi_2 =$ Mary is available after dinner.

The proposal: second part The precise shape of the property \mathcal{R} is of course determined by the context. Yet, let me assume that grammar imposes that suitable assignments to \mathcal{R} satisfy certain postulates. I submit the two following postulates (89). The postulate (89a) is the Gricean Maxim of Relevance, that says that an utterance must always be relevant. The postulate (89b) says that relevance is closed with respect to contextual equivalence, thus capturing the intuition that \mathcal{R} is a contextual variable and is thus sensitive to contextual information.

- (89) a. If Φ is uttered, then $\mathcal{R}(\Phi)$.
 b. If $\varphi \leftrightarrow_{w_{ck}} \psi$, then $\mathcal{R}(\varphi) = \mathcal{R}(\psi)$.

Before I move on, let me pause to note the parallelism between the theory of the exhaustivity operator sketched here and some versions of the theory of the distributivity operator, such as that of Schwarzschild (1996). Both the exhaustivity and the distributive operator are covert operators. Both of them are closely related to corresponding overt operators: ‘only’ for the case of the exhaustivity operator; ‘each’ for the case of the distributive operator. Both the exhaustivity and the distributive operator are relativized to a proper parameter supplied by the context: a relevance relation \mathcal{R} for the exhaustivity operator; a cover \mathcal{C} for the distributive operator. In both cases, the contextual parameters are constrained by grammatical postulates: the relevance property has to be closed under contextual equivalence; covers have to sum up to the entire domain. Both the exhaustivity and the distributivity operator can be assumed to be mandatorily present: lack of implicatures does not correspond to lack of the exhaustivity operator but to a proper choice of \mathcal{R} (so that the scalar alternatives are not relevant); analogously, a collective reading does not correspond to lack of the distributivity operator but to a proper choice of \mathcal{C} (a collective cover).

- (90) Analogies between the exhaustivity operator EXH and the distributivity operator DIST:
- both are covert operators. . .
 - . . . with an analogous overt operator;
 - both depend on a context-supplied parameter. . .
 - . . . whose possible values are constrained by grammatical axioms;
 - both can be assumed to be mandatorily present at LF.

I am now ready to show how the oddness of the sentences (60)-(67) considered at the beginning of this section is accounted for by means of the assumptions spelled out so far. Let me concentrate for instance on sentence (64a), repeated below in (91). As noted in chapter 1, Hawkins’ account (45) maintains that sentence (91) sounds odd because it triggers the scalar implicature that not all Italians come from a beautiful country, which mismatches with the piece of common knowledge that all Italians come from the same country. Here is how this piece of reasoning is formalized on the background on the assumptions introduced in this section.

- (91) #Some Italians come from a beautiful country.

By assumption (86), sentence (91) admits one and only one LF, namely Φ in (92), with a matrix exhaustivity operator whose prejacent I have called φ .

- (92) [Φ EXH [φ Some Italians come from a beautiful country]]

Let me stick with the classical assumption (93) that ‘some’, ‘all’ is a Horn scale. By definition (74), the set $\text{Alt}(\varphi)$ of scalar alternatives of the prejacent φ thus contains the alternative ψ obtained from φ by replacing ‘some’ with ‘all’.

- (93) $\text{Alt}(\varphi) = \{\psi = \text{All Italians come from a beautiful country}\}$

Of course, ψ logically entails φ .¹¹ But the opposite doesn't hold, since there is of course a world, not compatible with common knowledge \mathcal{W}_{ck}^1 , where Italians come from different countries and only some of them happen to come from one that is beautiful. Thus, the alternative ψ asymmetrically entails the prejacent φ if we do not take common knowledge into account. By virtue of the BH (82), this is enough to conclude that ψ is excludable given φ according to the definition (75) of the set of excludable alternatives, as stated in (94). Note the crucial role played by the BH (82): if the strengthened meaning were computed using entailment relative to common knowledge, then ψ could not have counted as an alternative excludable w.r.t. φ , since φ and ψ are equivalent given the piece of common knowledge that all Italians come from the same country.

$$(94) \text{ Excl}(\varphi) = \{\psi\}$$

By the slightly amended definition (87b) of the exhaustivity operator EXH, the proposition denoted by the odd sentence (91) is (95), dependent on the contextual relevance predicate \mathcal{R} .

$$(95) \text{ EXH}_{\mathcal{R}}(\varphi) = \varphi \wedge (\neg\psi \vee \neg\mathcal{R}(\psi)).$$

Consider any context where the piece of common knowledge holds that all Italians come from the same country. In any such context, the prejacent φ and its alternative ψ are contextually equivalent. Thus, they pattern alike with respect to relevance, by assumption (89b). Let me distinguish two cases. Consider first the case where the prejacent φ is not relevant. In this case, ψ is not relevant either. Hence, the strengthened meaning (95) boils down to just φ . Since the latter is not relevant by hypothesis, then it follows that the LF (92) uttered by the speaker is not relevant, contradicting assumption (89a). Thus, I only need to consider the case where φ is relevant. In this case, ψ is relevant too. Hence, the strengthened meaning (95) boils down to (96).

$$(96) \text{ EXH}_{\mathcal{R}}(\varphi) = \varphi \wedge \neg\psi$$

In conclusion, the LF (92) denotes the proposition $\varphi \wedge \neg\psi$ that “some but not all Italians come from a beautiful country” in any context where the piece of common knowledge holds that all Italians come from the same country. Of course, this proposition is a contradiction given common knowledge. The oddness of sentence (91) is thus predicted by the plausible linking condition (97), that I will call the MISMATCH HYPOTHESIS (henceforth: MH). As far as I can see, the account just presented straightforwardly extends to the other sentences (60)-(67). In Subection 3.1, I'll suggest that the oddness of sentence (2b), with the adverb ‘sometimes’ and the i-predicate ‘tall’, can be accounted for in exactly the same way.

$$(97) \text{ MISMATCHING HYPOTHESIS. If a sentence denotes a contradiction given common knowledge, then that sentence sounds odd.}$$

Before I move on, let me comment on assumption (97). If a given sentence admits a unique LF and if the proposition denoted by that LF is a contradiction given common knowledge, then the MH (97) says that that sentence sounds odd in the context considered. What if a sentence admits two LFs and only one of them denotes a contextually contradictory proposition? I assume that, if a sentence is ambiguous between two LFs and one of them is pragmatically deviant, then that sentence is not perceived as pragmatically deviant but the ambiguity is lost, namely the pragmatically fine LF can “cover up” the pragmatically deviant LF to the point that the latter is not accessible anymore. Thus in particular, if a sentence might correspond to two LFs one of which denotes a contextually contradictory proposition, then (97) predicts that that sentence sounds fine but that the meaning corresponding to the guilty LF is not available.

Remark Let me close this section by addressing a recent debate concerning the relationship between the grammatical nature of scalar implicatures and their mandatoriness. It is sometimes claimed in the literature that, by virtue of their very nature, grammatical theories of scalar implicatures necessarily predict implicatures to be mandatory, or default, or generated automatically and mechanically

¹¹ This is strictly speaking not true, because of possible worlds where there are no Italians, and thus the restrictor of the existential and universal operators is empty.

without any need for contextual triggering. For instance, Geurts (2009, p. 58) writes: “There is another distinction that I cannot afford to ignore, viz. between DEFAULTIST and NONCIST varieties of localism.¹² If you are a defaultist, you are committed to the view that scalar expressions give rise to upper-bounding inferences as a matter of course; it is what will happen normally. [...] Chierchia and Levinson are both defaultists in this sense of the word. Other authors adopt the position that scalar inferences are entirely dependent on the context; they are noncists. [...] Methodologically speaking, noncist versions of localism are disappointingly weak in the sense that they merely predict that local inferences may or may not occur. Clearly, this type of theory needs to be complemented by an account of why and when scalar inferences arise. As far as I am aware, there is no such theory on the market.” And Geurts and Pouscoulous (2009a, p. 23), slightly refining Geurts and Pouscoulous (2009b), write: “Thus far, we have been concerned with what we have dubbed ‘mainstream conventionalism’¹³, a label for a spectrum of theories all of which predict that local strengthening should be preferred [...]. We presented data showing, arguably, that the predictions of mainstream conventionalism are wrong. What to do? [...] [One] option is to leave the mainstream without giving up on conventionalism: stick to your favourite [conventionalist] brand of scalar implicatures, which will duly generate a batch of interpretations for any sentence that contains scalar expressions, but refuse to make predictions about which construal is the preferred one. Leave it to pragmatics. To the best of our knowledge, nobody has come forward to advocate such a minimalist take on conventionalism, as yet, and we are tempted to say that this is just as well, since this view is too weak to be taken seriously.” In this section, I have argued in favor of the three claims (98), thus defending precisely this view that Geurts and Pouscoulous think cannot be taken seriously.

- (98)
- a. Grammatical theories of scalar implicatures *per se* do not make any prediction on the context sensitivity of scalar implicatures
 - b. Grammatical theories of scalar implicatures are very well compatible with the context sensitivity of scalar implicatures
 - c. Grammatical theories of scalar implicatures do not owe an account of the fine grained contextual dependence of scalar implicatures

Let me review my arguments for (98). The grammatical theory of scalar implicatures that I have adopted here from Fox (2007a) really just says that, *if* an implicature is generated, that must be signaled in the LF of the sentence, by appending a suitable operator. This proposal *per se* does not make any prediction concerning the context sensitivity vs. mandatoriness of scalar implicatures. I have suggested a way to capture the observed contextual sensitivity of scalar implicatures within this grammatical theory of implicatures. My proposal is completely straightforward: implicature cancellation boils down to the very well known mechanism of contextual domain restriction. More in detail, I have made the following suggestion: the exhaustivity operator is a universal quantifier; its restrictor is partially contextually determined, just like it is the case for any other universal quantifier; whether an implicature is derived or not, that depends on whether the corresponding alternative is relevant (and thus makes it to the restrictor of the exhaustivity operator) or not. A theory of implicatures does not owe a theory of relevance, just as a theory of distributivity does not owe a theory of covers. The latter point is made very explicitly in Schwarzschild (1996, p. 92): “On the view exposed here, the truth conditions for sentences with plural arguments are often determined in part by the assignment to a free variable over coverings [...] of the domain. We have said that the source of this assignment is pragmatic. Can we say more? The question of what makes a partitioning of the domain salient in the discourse bears some resemblance to the question of what makes the antecedent for a pronoun salient. In many instances there are linguistic clues, [...] but arriving at a complete answer surely involves other branches of cognitive science. Such is the case for domain partitioning as well. How we divide up our visual space for example is relevant here and yet that is a question which is properly a matter yet to be settled by experts on vision.”

¹²My understanding is that what Geurts means by “localism” is what I mean by “grammatical theories” of scalar implicatures.

¹³Again, my understanding is that what Geurts means by “conventionalism” is what I mean by “grammatical theories” of scalar implicatures.

2.2 Extension to cases with multiple alternatives

In section 2.1, I have considered a first set of odd sentences (60a)-(67a). I have argued that those sentences sound odd because they come with a logically stronger alternative ψ that yields a blind scalar implicature $\neg\psi$ that mismatches with the relevant piece of common knowledge. In this section, I consider some more complicated cases of oddness, where mismatch with common knowledge comes about not from a *single* implicature $\neg\psi$ but from the conjunction $\neg\psi_1 \wedge \neg\psi_2 \wedge \neg\psi_3 \dots$ of *multiple* implicatures $\neg\psi_1, \neg\psi_2, \neg\psi_3, \dots$. The cases considered in this section will turn out to be important for my discussion of restrictions on bare plural subjects of i-predicates in chapter 3.

The facts Fronting of the universal quantifier ‘every day’ in (99a) forces it to take wide scope over the existential quantifier ‘a fireman’. Sentence (99b) has no fronting and thus admits also the reading with wide scope existential. Since the reading with narrow scope existential admitted by sentence (99a) is entailed by the reading with wide scope existential admitted by sentence (99b), then we might have expected sentence (99a) to sound just as fine as sentence (99b) in the context considered. We are thus faced with the puzzle of accounting for the contrast in (99).¹⁴

- (99) *Context*: a competition lasted five days, Monday through Friday; both John and Bill know that the same guy x won on each of the five days:

	M	T	W	T	F
<i>winner</i> :	x	x	x	x	x

John wants to know more about this amazing guy and thus asks Bill for more information; Bill provides the following information:

- #On every day, a/some fireman won.
- A/Some fireman won on every day.

Analogously, the bound pronoun in (100a) forces the universal quantifier ‘every friend’ to have wide scope over the indefinite ‘an american girl’. Even with this scope configuration, the plain meaning of sentence (100a) is of course compatible with the boys having all fallen in love with the same girl they had met at the party, and thus the continuation ‘this girl...’ should be felicitous, just as in the case of sentence (100b), which admits the opposite scope configuration, because does not contain the bound pronoun. Thus, we are faced with the puzzle of accounting for the contrast in (100).

- (100) a. Every friend of mine loves an American girl he had met at the party.
#This girl is really beautiful.
- b. Every friend of mine loves an American girl.
This girl is really beautiful.

Consider next the following variant (101) of the example (99) considered above. Sentence (101a) feels odd in the context considered. In sentence (101b), the definite object ‘the running competition’ has been replaced by the wide scope universal object ‘every competition’ and the sentence feels fine in the context considered. Whatever accounts for the oddness of sentence (101a), must be able to derive the fact that the sentence turns fine once embedded under a further universal quantifier, as in (101b).

- (101) *Context*: a competition lasted for five days, Monday to Friday; each day, three challenges are held: swimming, running and jumping; both John and Bill know that the same guy x won the swimming competition on all five days, the same guy y won the running competition on all five days and the same guy z won the jumping competition on all five days:

	M	T	W	T	F
<i>swimming</i> :	x	x	x	x	x
<i>running</i> :	y	y	y	y	y
<i>jumping</i> :	z	z	z	z	z

¹⁴The example in (99) is a slight variant an example from Percus (2001). I discuss Percus’ original example in the last Remark of this section.

John wants to know more about these amazing guys, and thus asks Bill for more information; Bill replies as follows:

- a. #Every day, a/some fireman won the running competition.
- b. Every day, for every competition, a/some fireman won.

Analogously, consider the following variant (102) of the example (100) considered above. It is hard to accommodate the continuation ‘the woman the boys were introduced to by Mary’ in the case of sentence (102a). In sentence (102b), ‘Mary’ has been replaced by the universal quantifier ‘each of Mary, Ann and Sue’ and the continuation ‘the woman the boys were introduced to by Mary’ feels more straightforward.

- (102) a. Mary introduced every boy to a woman he finds attractive.
#The woman the boys were introduced to by Mary is really beautiful.
- b. Mary, Ann and Sue each introduced every boy to a woman he finds attractive.
The woman the boys were introduced to by Mary is really beautiful.

In the rest of this subsection, I concentrate on (99) and (101), the case of (100) and (102) being formally parallel. I will argue that these puzzles are solved by the conjunction $\neg\psi_1 \wedge \neg\psi_2 \wedge \dots$ of multiple blind implicatures. Let me thus pause on some preliminaries on the computation of implicatures when the set of excludable alternatives is not a singleton, as it was the case for the examples considered in section 2.1.

Preliminaries Horn-scales need not really be scales, namely sets *linearly* ordered w.r.t. entailment, but rather just *partially* ordered sets. The main such example is the case of sentences containing disjunction, such as φ in (103a): Sauerland (2004c) suggests that the set of scalar alternatives $\mathcal{Alt}(\varphi)$ is the one in (103b), which contains, besides the corresponding sentence ψ_{and} with conjunction, also the left and the right conjuncts ψ_L and ψ_R .

- (103) a. $\varphi =$ Kai had peas or broccoli.
- b. $\mathcal{Alt}(\varphi) = \left\{ \begin{array}{l} \varphi = \text{Kai had peas or broccoli} \\ \psi_L = \text{Kai had peas} \\ \psi_R = \text{Kai had broccoli} \\ \psi_{\text{and}} = \text{Kai had peas and broccoli} \end{array} \right\}$

A natural generalization of assumption (103) concerns sentences containing an existential quantifier, such as sentence φ in (104a): I assume that its set of scalar alternatives $\mathcal{Alt}(\varphi)$ is the one in (104b), which contains, besides the corresponding sentence ψ_{all} with the universal quantifier, also the three alternatives ψ_{John} , ψ_{Bill} and ψ_{Tom} obtained by replacing the existential quantifier ‘some of his brothers’ with the definite description ‘his brother such and such’, where ‘such and such’ can be for instance ‘called John/Bill/Tom’.¹⁵

¹⁵The obvious *formal* correspondence between (104) and (103) is spelled out in (i).

- (i) a. existential quantifier \leftrightarrow disjunction;
- b. universal quantifier \leftrightarrow conjunction;
- c. definite descriptions \leftrightarrow two disjuncts.

Sentence (iia) triggers a very strong IGNORANCE INFERENCE that the speaker does not know which one of his two sisters Mary and Sue does John ate. Sauerland (2004c) suggests that this ignorance inference is evidence for his assumption that the two disjuncts are scalar alternatives of disjunction. In order to support my parallel assumption for existential quantifiers, let me point out that sentence (iib) seems to trigger the same ignorance inference.

- (ii) a. John hates Mary or Sue.
- b. John hates one of his two sisters.

Sauerland derives (103b) from the assumption that disjunction is a Horn-mate of the two operators LEFT = $\lambda p . \lambda q . q$ and RIGHT = $\lambda p . \lambda q . p$. In this way, (103b) fits the definition (74) of the set of scalar alternatives. In any case, (103b) is compatible with the definition of scalar alternatives of Katzir (2008), as the set of those LFs that can be obtained from the target LF by deletion or lexical substitution. It is not completely clear to me how to make my variant (104) compatible with the definition (74) of the set of scalar alternatives or with Katzir’s theory of scalar alternatives. One possible way to go might

- (104) a. $\varphi = \text{Kai met some of his three brothers.}$
- b. $\text{Alt}(\varphi) = \left\{ \begin{array}{l} \varphi = \text{Kai met some of his three brothers} \\ \psi_{\text{John}} = \text{Kai met his brother John} \\ \psi_{\text{Bill}} = \text{Kai met his brother Bill} \\ \psi_{\text{Tom}} = \text{Kai met his brother Tom} \\ \psi_{\text{all}} = \text{Kai met all of his brothers} \end{array} \right\}$

Once we allow for sets of scalar alternatives such as (103) and (104), the simple definition (75) of the set of excludable alternatives runs into a well-known problem; let me illustrate the problem with the case in (104). Each of the three scalar alternatives ψ_{John} , ψ_{Bill} and ψ_{Tom} in (104b) asymmetrically entails φ in (104a). Hence, if the strengthened meaning $\text{EXH}(\varphi)$ of φ was computed using the classical definition (75) of the set $\text{Excl}(\varphi)$ of excludable alternatives, we would get (105). Yet, this strengthened meaning is a logical contradiction. Thus, the classical definition (75), coupled with the assumption (104) on scalar alternatives, fails to derive the intended strengthened meaning of sentence φ , namely that Kai met some but not all of his brothers.

$$(105) \quad \text{EXH}(\varphi) = \varphi \wedge \neg\psi_{\text{John}} \wedge \neg\psi_{\text{Bill}} \wedge \neg\psi_{\text{Tom}} \wedge \neg\psi_{\text{all}}$$

Fox (2007a), building on Sauerland (2004c), solves this problem by replacing the classical definition (75) of the set of excludable alternatives with the alternative definition (106). The intuition behind (106) is the following: the strengthened meaning $\text{EXH}(\varphi)$ asserts φ and excludes as many scalar alternatives ψ 's as can be excluded in a non-arbitrary way without getting a contradiction.

- (106) a. A subset $X = \{\psi_1, \psi_2, \dots\}$ of the set $\text{Alt}(\varphi)$ is called an **CONSISTENTLY EXCLUDABLE** subset w.r.t. φ iff $\varphi \wedge \neg\psi_1 \wedge \neg\psi_2 \wedge \dots$ is not a contradiction.
- b. A subset $X \subseteq \text{Alt}(\varphi)$ is called a **MAXIMAL** consistently excludable subset w.r.t. φ iff there are no consistently excludable supersets of X in $\text{Alt}(\varphi)$.
- c. The set $\text{Excl}(\varphi)$ of **EXCLUDABLE** alternatives w.r.t. φ is the intersection of all maximal consistently excludable subsets of $\text{Alt}(\varphi)$ w.r.t. φ .

Here is how this new definition solves our problem: according to clause (106b), there are three maximal subsets of excludable alternatives: $\{\psi_{\text{John}}, \psi_{\text{Bill}}, \psi_{\text{all}}\}$, $\{\psi_{\text{John}}, \psi_{\text{Tom}}, \psi_{\text{all}}\}$ and $\{\psi_{\text{Bill}}, \psi_{\text{Tom}}, \psi_{\text{all}}\}$; according to clause (106c), $\text{Excl}(\varphi)$ is their intersection, namely $\text{Excl}(\varphi) = \{\psi_{\text{all}}\}$; hence, we obtain $\text{EXH}(\varphi) = \varphi \wedge \neg\psi_{\text{all}}$, which is the right result.

First part of the account I want to suggest an account for the oddness of sentence (99a) completely analogous to that suggested in section 2.1 for the oddness of sentences (60)-(67). The idea of this account is presented in (107), repeated from (52).

- (107) a. On the one hand, an utterance of (99a), with narrow scope indefinite, triggers the scalar implicature that the logically stronger alternative (99b), with wide scope indefinite, is false.
- b. On the other hand, common knowledge entails that the same guy always won, hence (99a) cannot be true without (99b) being true too.
- c. In conclusion, the oddness of sentence (99a) follows from the mismatch between the scalar implicature in (107a) and common knowledge in (107b).

The crucial problem here is how to formalize step (107a). The two sentences (99a) and (99b) differ for the relative scope of the universal and the existential quantifier. Thus, sentence (99b) cannot be

be as follows. Following for instance Westerthal (1984) and von Stechow (1994), let me assume that the LF of 'some NP' is something like (ii). Here, \mathbf{C} is the name of a contextually determined free variable of type $\langle e, t \rangle$. The constituent $[\text{NP } \mathbf{C}]$ is interpreted by predicate modification.

- (ii) [some [NP \mathbf{C}]]

I could then try to make (104) compatible with the definition (74) of the set of scalar alternatives and with Katzir's theory of scalar alternatives by assuming that the variable \mathbf{C} is a Horn mate of / can be replaced by any other variable \mathbf{C}' and that one of these variables \mathbf{C}' denotes a property "such and such" whose intersection with NP is a singleton. I leave the issue open.

derived from sentence (99a) just by replacing scalar items with Horn-mates. The current definition (74) of the set of scalar alternatives thus straightforwardly predicts (108). Since (99b) is not a scalar alternative, then of course it cannot trigger any implicature and thus step (107a) fails. One might be tempted to take this difficulty at face value, and conclude that the current definition (74) of the set of scalar alternatives is too strict, and that it should be replaced with an alternative definition that makes the set of scalar alternatives closed with respect to reshuffling of the quantifiers. I will discuss this option in the last Remark at the end of this section.

(108) Sentence (99b) is *not* a scalar alternative of sentence (99a).

For the time being, I will stick with the current definition (74) of the set of scalar alternatives and suggest that the implicature needed for the first step (107a) can be derived in an *indirect* way, that circumvents the problem in (108). Here are the details. The LF of the odd sentence (99a) is the one in (109a). This LF corresponds to the truth conditions φ in (109b), namely that “for each day t of the competition, there was some fireman x such that x won on t .” Note that ‘each day’ (i.e. \forall_t) has scope over ‘some fireman’ (i.e. \exists_x) in (109b): as anticipated, this will turn out to be the culprit for the oddness of the sentence.

- (109) a. [[each day] [λt [[some fireman] [λx [x won in t]]]]].
 b. $\varphi = \lambda w . \forall_t [[\text{day}]^w(t) \rightarrow [\exists_x [[\text{fireman}]^w(x) \wedge [\text{win}]^w(x, t)]]]$.

As already assumed in (104) and stated in (110), let me assume that the existential quantifier ‘a/some fireman’ and the definite description ‘the fireman such and such’ are Horn-mates, for any choices of ‘such and such’. Here and henceforth, I’ll use ‘the fireman P ’ as a shorthand for ‘the fireman such and such’ and I’ll use d_P^w as a shorthand for the individual [[the fireman P]]^w.

(110) ⟨‘a/some fireman’, ‘the fireman P ’⟩ is a Horn-scale.

Consider now the LF (111a), which is obtained from the LF (109a) by replacing ‘some fireman’ by ‘the fireman such and such’ (abbreviated as ‘the fireman P ’). Since these two scalar items are Horn-mates by assumption (110), then (111a) is a scalar alternative of (109a), according to the standard definition (74) of the set of scalar alternatives. The truth conditions of the LF (111a) are ψ_P in (111b), namely that “the fireman such and such won on every day t ”.

- (111) a. [[each day] [λt [[the fireman P] [λx [x won in t]]]]].
 b. $\psi_P = \lambda w . \forall_t [[\text{day}]^w(t) \rightarrow [\text{win}]^w([\text{the fireman } P]^w, t)]$.

Let ψ be the truth conditions (112) of the fine sentence (99b), construed with the indefinite over firemen \exists_x having wide scope over the universal operator \forall_t over the days of the competition.

- (112) $\psi = \lambda w . \exists_x [[\text{fireman}]^w(x) \wedge \forall_t [[\text{day}]^w(t) \rightarrow [\text{win}]^w(x, t)]]]$.

The logical equivalence in (113) holds:¹⁶ the conjunction of the negation of all the alternatives ψ_P in (111) is equivalent to the negation of the truth condition ψ of the fine sentence (99b), i.e. to the statement that “no single fireman won on every day of the competition.” The equivalence (113) is my strategy to circumvent the problem raised in (108).

$$(113) \quad \bigwedge_P \neg \psi_P = \neg \psi$$

Let me now argue that the strengthened meaning $\text{EXH}(\varphi)$ of φ in (109b) is the one in (114), i.e. that each ψ_P is excludable w.r.t. φ . Given the logical equivalence in (113), it is sufficient to show that $\varphi \wedge \neg \psi$ is not a logical contradiction. And of course it is not, since there are possible worlds where different firemen won on different days, where thus φ is true but ψ is false.

$$(114) \quad \text{EXH}(\varphi) = \varphi \wedge \underbrace{\bigwedge_P \neg \psi_P}_{\neg \psi}$$

¹⁶In fact, $\bigwedge_P \neg \psi_P = \neg \bigvee_P \psi_P$, by De Morgan laws; and furthermore $\bigvee_P \psi_P = \psi$, since $\{d_P^w \mid P\} = [\text{fireman}]^w$.

Note nonetheless that $\text{EXH}(\varphi) = \varphi \wedge \neg\psi$ is a contradiction with respect to the piece of common knowledge that a single person won on every day. The oddness of (99a) thus follows from the MH (97). Note again the crucial role played by the BH (82): if the strengthened meaning were computed taking common knowledge into account, then the entire set of the ψ_P 's would not be excludable and thus no mismatch would arise. In Section 3, I'll argue that the lack of the existential reading of bare plural subjects of i-predicates illustrated in (7) can be accounted for in the same way as I have accounted here for the oddness of sentence (99a), building on the fact that existential bare plurals are known to always take the narrowest possible scope.

Second part of the account Let me argue that the contrast in (101) follows straightforwardly from the proposal just presented. The oddness of sentence (101a) can be accounted for in the same way as that of sentence (99a) was accounted for above. Let me now quickly repeat the reasoning above for the case of sentence (101b) in order to make sure that it does *not* predict this sentence to sound odd. The truth conditions of sentence (101b) are φ in (115), where \forall_t stands for 'every day of the week t ', \forall_y stands for 'every competition y ' and \exists_x stands for 'there is a fireman x '.

$$(115) \quad \varphi = \lambda w . \forall_y [\forall_t [\exists_x [\text{win}]^w(x, y, t)]]$$

Consider the alternative ψ_P in (116) obtained by replacing the indefinite 'a/some fireman' with the definite 'the fireman P ', whose denotation is as usual indicated as d_P^w .

$$(116) \quad \psi_P = \lambda w . \forall_y [\forall_t [\text{win}]^w(d_P^w, y, t)]$$

Let ψ be as in (117). Note the crucial fact that the existential quantifier over firemen \exists_x in the formula ψ has wide scope not only over the universal quantifier over days \forall_t but also over the universal quantifier over competitions \forall_y .

$$(117) \quad \psi = \lambda w . \exists_x [\forall_y [\forall_t [\text{win}]^w(x, y, t)]]$$

The logical equivalence in (118) holds:¹⁷ the conjunction of the negation of all the alternatives ψ_P in (116) is equivalent to the negation of ψ in (117), namely to the statement that "there is no fireman who won every competition on every day."

$$(118) \quad \bigwedge_P \neg\psi_P = \neg\psi$$

Thus, the (blind) strengthened meaning $\text{EXH}(\varphi)$ of φ boils down to (119), which says that "every day, for every competition, there is a fireman who won and that it is false that there is a fireman who won every competition on every day." Despite the fact that the strengthened meaning in (119) is formally analogous to that in (114), the former is *not* a contradiction given common knowledge. In fact, since the existential quantifier of ψ in (117) has wide scope also over the universal quantifier over competitions and since it does not follow from common knowledge that a single guy won all three of the competitions, then the negation of ψ is consistent with φ relative to the common knowledge.

$$(119) \quad \text{EXH}(\varphi) = \varphi \wedge \neg\psi$$

In Section 3, I'll argue that Fox's surprising effect (8), that a bare plural subject of an i-predicate does admit the existential reading when it occurs in the scope of a universal quantifier, can be accounted for in the same way as I have accounted here for the contrast in (101). In the rest of this section, I qualify the account just presented with a number of remarks.

First remark: the proper definition of the set of scalar alternatives To secure my account for the contrast in (101), I need to discuss a technical issue concerning the proper definition of the set of scalar alternatives. Consider ϕ_P in (120), for any choice of the property P . This formula ϕ_P is obtained from formula φ in (115) by means of two replacements: the universal quantifier over competitions \forall_y is replaced by the existential quantifier \exists_y ; and the existential quantifier over firemen \exists_x is replaced by the individual d_P^w denoted by the definite description 'the fireman P '.

¹⁷The reasoning is identical to that in footnote 16.

$$(120) \quad \phi_P = \lambda w . \exists y [\forall t [[\text{win}]^w(d_P^w, y, t)]] .$$

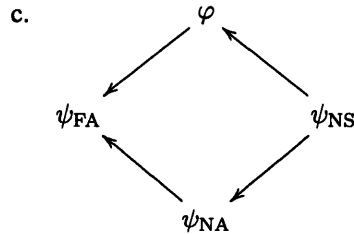
The logical equivalence in (121) holds: the conjunction of the negation of all the ϕ_P 's in (120) is equivalent to the negation of the proposition ϕ that “there exists a fireman x and a competition y such that x won y on every day”. Of course, $\varphi \wedge \neg\phi$ is not a logic contradiction. Yet, it is indeed a contradiction given the piece of common knowledge considered in (101). Thus, if the ϕ_P 's in (120) were scalar alternatives of φ in (115), then the blind strengthened meaning of φ would entail that $\varphi \wedge \neg\phi$ and would therefore contradict common knowledge, contrary to the fact that sentence (101b) is fine in the context considered. Let me show that a solution to this problem follows from a more careful definition of the set of scalar alternatives, independently needed.

$$(121) \quad \bigwedge_P \neg\phi_P = \neg\phi, \quad \text{where } \phi = \lambda w . \exists x [\exists y [\forall t [[\text{win}]^w(x, y, t)]]] .$$

The new definition (106) of the set $\mathcal{E}xcl(\varphi)$ of alternatives excludable w.r.t. φ differs from the standard definition (75) because it makes use of the notion of contradictoriness instead of the notion of asymmetric entailment. Precisely for this reason, this new definition (106) leads to a slight problem when coupled with the standard definition (74) of the set $\mathcal{A}lt(\varphi)$ of the scalar alternatives of a sentence φ . The problem is completely general, and arises whenever a non-maximal scalar item is embedded under another non-maximal scalar item; let me illustrate it with the case of sentence φ in (122a). According to the standard definition (74), the set $\mathcal{A}lt(\varphi)$ of scalar alternatives in the case of (122a) contains the three alternatives ψ_{FA} , ψ_{NS} and ψ_{NA} in (122b), under the standard assumption that ‘some’, ‘all’ and ‘few’, ‘none’ are Horn-scales.¹⁸ These alternatives are ordered with respect to (asymmetric) entailment as in (122c).

(122) a. $\varphi =$ Few boys did any of the readings.

$$b. \quad \mathcal{A}lt(\varphi) = \left\{ \begin{array}{l} \psi_{FA} = \text{Few boys did all of the readings.} \\ \psi_{NS} = \text{No boys did some of the readings.} \\ \psi_{NA} = \text{No boys did all of the readings.} \end{array} \right\}$$



According to the standard definition (75) of the set $\mathcal{E}xcl(\varphi)$ of excludable alternatives, only asymmetrically entailing alternatives are excludable; hence, only ψ_{NS} is predicted to be excludable in this case, and the right strengthened meaning is thus derived. Things are very different if we switch to the new definition (106) of the set $\mathcal{E}xcl(\varphi)$ of excludable alternatives. Of course, ψ_{FA} is not excludable given φ , since ψ_{FA} is entailed by φ . Thus, the set of excludable alternatives according to (106) is $\mathcal{E}xcl(\varphi) = \{\psi_{NS}, \psi_{NA}\}$, since these two alternatives ψ_{NS} and ψ_{NA} can be both negated consistently with φ . In conclusion, the strengthened meaning of sentence φ is predicted to be (123), namely “few boys did some of the readings and some boy did some of the readings and some boy did all of the readings.” This result is wrong: an utterance of φ does not in any way commit to the existence of a boy who has done all of the readings.

$$(123) \quad \text{EXH}(\varphi) = \varphi \wedge \neg\psi_{NS} \wedge \neg\psi_{NA} .$$

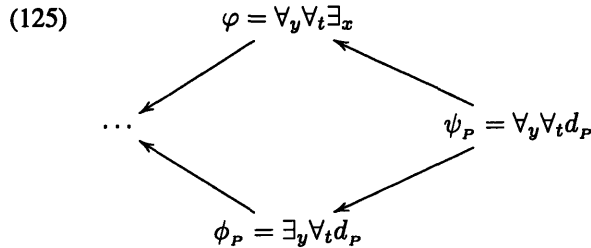
In order to avoid this problem, we need to modify the standard definition (74) of the set $\mathcal{A}lt(\varphi)$ of scalar alternatives in such a way to get ψ_{NA} out of it. Fox (2007a, footnote 35) suggests to replace the standard definition with the following alternative definition (124). The intuition behind (124) is the following: each scalar alternative ψ of φ is obtained from φ by a sequence of steps; at each step, only one scalar item can be replaced; and a replacement is licit iff it leads to something not weaker.

¹⁸Here, I ignore the fact that ‘some’ is a PPI.

(124) The set $Alt(\varphi)$ is recursively defined as follows:

- a. $\varphi \in Alt(\varphi)$;
- b. $\psi \in Alt(\varphi)$ iff there is $\phi \in Alt(\varphi)$ such that ψ is not weaker¹⁹ than ϕ and furthermore ψ is obtained from ϕ by replacing a single scalar item in ϕ with a Horn-mate.

Given a diagram such as (122c), the scalar alternatives of a sentence φ according to (124) can be described as those nodes ψ 's of the diagram such that it is possible to walk from ψ to φ on left-oriented arrows only. Thus, the new definition (124) predicts the offending ψ_{NA} not to be a scalar alternative of φ , since there is no way to reach the target φ starting from ψ_{NA} by walking on left-oriented arrows only in the diagram in (122c). The problem posed by (122a) is thus solved by the new definition (124).²⁰ And the problem of the offending ϕ_P 's in (120) is solved as well. In fact, it is not possible to walk from the offending ϕ_P to the target φ along left-oriented arrows only in the diagram in (125).



For the sake of simplicity, in the body of the paper, I will stick to the simpler, standard definition (74) of the set of scalar alternatives, and point out in footnotes that my proposal is fully compatible with the refined definition (124).

Second remark: the issue of relevance The preceding discussion of the oddness of sentence (99a) has overlooked one crucial issue: why are the alternatives ψ_P in (111) *relevant*? why is the corresponding implicature mandatory? Let me make the problem explicit. In section 2.1, I have started out with the simple definition (87a) of the exhaustivity operator. For the simple case of the odd sentences (60a)-(67a) considered in section 2.1, this definition yields the strengthened meaning in (126a), that contains the negation of the unique alternative ψ . For the more complicated case of the odd sentence (99a) considered in this section, this definition yields the strengthened meaning (114), repeated in (126b), in terms of the many alternatives ψ_P in (111). In both cases, the strengthened meaning contradicts common knowledge, and thus oddness follows.

- (126) a. $EXH(\varphi) = \varphi \wedge \neg\psi$.
- b. $EXH(\varphi) = \varphi \wedge \underbrace{\bigwedge_P \neg\psi_P}_{\neg\psi}$.

Yet, in order to account for the context dependence of implicatures, in section 2.1 I have replaced the original definition (87a) with the variant (87b), dependent on the relevance predicate \mathcal{R} . Once this slightly modified definition is in place, the two strengthened meanings in (126) need to be replaced

¹⁹Of course, given the BH, I take “non weaker” to mean “logically non weaker”.

²⁰Note that, while the alternative ψ_{NS} asymmetrically entails φ , the alternative ψ_{NA} does not, since it is logically independent of φ . Thus, the problem posed by sentence (122a) could have been solved also by other conceivable modifications of the original definition (74) of the set of scalar alternatives, such as those in (ia) or (ib).

- (i) The set $Alt(\varphi)$ contains all and only those ψ 's that can be obtained from φ by replacing scalar items in φ with Horn-mates and that furthermore ...
 - a. ... asymmetrically entail φ .
 - b. ... are not independent of φ .

Fox (2007a) argues against both (ia) and (ib) and in favor of (124), by noting that (ia) and (ib) make wrong predictions for the case of scalar items embedded under non-monotone operators.

by the slight variants in (127). This replacement turned out to have no consequences for the simple case (127a) of the odd sentences (60a)-(67a) considered in section 2.1. In fact, the unique alternative ψ is contextually equivalent to the prejacent φ ; hence, it is necessarily relevant, by assumptions (89); thus, (127a) turns out to be equivalent to (126a).

$$(127) \quad \text{a. } \text{EXH}(\varphi) = \varphi \wedge \left(\neg\psi \vee \neg\mathcal{R}(\psi) \right).$$

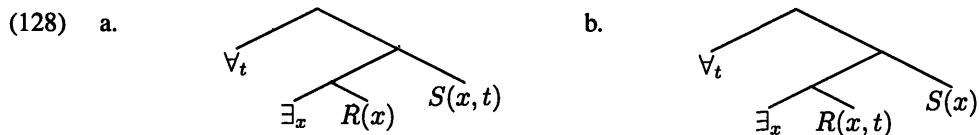
$$\text{b. } \text{EXH}(\varphi) = \varphi \wedge \bigwedge_{\psi_P} \left(\neg\psi_P \vee \neg\mathcal{R}(\psi_P) \right).$$

Unfortunately, I see no way to extend this line of reasoning so as to get (127b) to turn out equivalent to (126b). Assumption (89) on relevance does not in any way help in this case: it just ensures that $\neg\psi = \bigwedge_P \psi_P$ is relevant (since relevance is closed under negation and ψ is contextually equivalent to the prejacent φ); but in no way does it ensure that the single alternatives ψ_P are relevant too. Nor do I see any way to get out of trouble by adding plausible axioms on the relevance predicate \mathcal{R} . What I would really need to get out of trouble would be to derive the strengthened meaning (127b') instead of the one in (127b). Then, I could reason just as for the case of (127a), namely that ψ , or equivalently $\neg\psi$, must be relevant because contextually equivalent to the prejacent φ , so that (127b') effectively reduces to (126b).

$$(127) \quad \text{b'} \quad \text{EXH}(\varphi) = \varphi \wedge \left(\underbrace{\left(\bigwedge_{\psi_P} \neg\psi_P \right)}_{\neg\psi} \vee \neg\mathcal{R} \left(\underbrace{\bigwedge_{\psi_P} \neg\psi_P}_{\neg\psi} \right) \right).$$

The nature of the problem can be restated as follows. To overcome the problem (108), I have broken up the proposition ψ in (112) into the many small pieces ψ_P in (111) and I have used the universal quantification built into the exhaustivity operator to put the pieces back together, as in (126b). Truth conditions do not see the patch work, because they only look at the final result, after the exhaustivity operator has put the pieces back together. According to (127b), relevance unfortunately does see the patch work, because it kicks in before the exhaustivity operator has put the pieces back together. In order to get out of trouble, I would need to find a way to let relevance kicks in only after the pieces have been put back together, as in (127b').

Third remark: Percus' original sentence Sentence (99a) has the abstract form (128a). It has a wide scope universal quantifier \forall_t and a narrow scope existential quantifier \exists_x . The existential quantifier takes of course two arguments, namely a restrictor R and a nuclear scope S . The universal quantifier binds a variable t that sits inside the nuclear scope S of the existential quantifier. It is of course interesting to consider the variant (128b) of the scheme (128a), that only differs because of the fact that the universal quantifier binds a variable t that sits inside the restrictive clause R of the existential operator, rather than inside its nuclear scope.



Percus (2001) notes that (129a) sounds odd in the context considered. Plausibly, the adverb 'always' in (129a) must bind the time argument of the predicate 'finished'. The variant in (130) is completely analogous, only with an indefinite instead of a definite. Thus, these examples make the point that the case in (128b) behaves analogously to the one in (128a) as far as oddness is concerned.

- (129) *Context:* John and Bill were both present at a series of exams, which took place from Monday through Friday. Both of them saw that the same student finished first each time. John wants to know about the nationality of this surprising guy. Bill replies as follows:
- a. #The student who finished first was *always* Swedish.
 - b. The student who *always* finished first was Swedish.

- (130) *Context*: as in (129), with the only difference that the same three people finished first each time, all at the same time.
- #One of the students who finished first was *always* Swedish.
 - One of the students who *always* finished first was Swedish.

Unfortunately, the account that I have proposed here for the case of (128a) does not seem to extend to the case of (128b): I do not want to say that (129b) is an alternative to (129a) or that (130b) is an alternative to (130a); but I do not see any *indirect* way to obtain the desired mismatching implicature in this case.²¹

Forth remark: can we construct alternatives by reshuffling operators? To account for the oddness of sentence (99a), schematized as φ in (131a), I want its strengthened meaning to entail the negation of the corresponding sentence (99b) with the opposite scope configuration, schematized as ψ in (131b). Since neither the standard definition (74) of the set of scalar alternatives nor the revised definition (124) considered above allow for reshuffling of the operators, then ψ in (131b) is predicted not to be a scalar alternative of φ in (131a), as stated in (108).²² I have circumvented

²¹Let me comment on the account for the oddness of sentence (129a) suggested in Percus (2001). He presents the intuitive idea as follows: “A speaker who utters [(129a)] in a scenario like [the one considered here] [...] is quantifying over more things than he needs in order to make his point. [...] Specifically, in using the sentence [(129a)] he is stating that all members of {Monday, ..., Friday} are such that the student who finished first on that day was Swedish. But he knows that we could have drawn this conclusion if he had informed us that all the times in a smaller set have this character. On our [...] scenario, one such smaller set is {Tuesday}.” Percus formalizes this intuition by means of the pragmatic principle in (i). Informally, (i) states a preference for quantifying over the smallest possible domain (at least in the case of universal quantifiers). The oddness of sentence (129a) would thus follow from the fact that it violates the principle in (i). The account would extend from the case schematized in (128b) considered by Percus to the case schematized in (128a) that I have considered in this section.

- (i) Do not use the LF $\varphi = \forall_x[\alpha(x)][\beta(x)]$ if there is an LF $\psi = \forall_x[\alpha'(x)][\beta(x)]$ such that $\alpha' \subsetneq \alpha$ and furthermore “it follows from what the parties to a conversation are taking for granted about the actual world that φ holds of the actual world as long as ψ holds of it.”

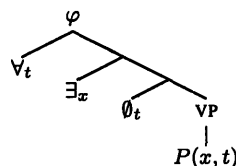
Yet, the principle in (i) is rather *ad hoc*. Furthermore, its application to (129) is actually doubtful, since the sentence does not make explicit the restrictor of ‘always’. Finally, this principle makes wrong predictions in cases such as (ii), which is a slight variant of example (62) from Chemla (2006).

- (ii) *Context*: Professor Smith always gives the same grade to all of his students. John wants to know the grade that professor Smith assigned to each of his students last semester. Bill provides the following information:
- φ = Last semester, all/the students of prof. Smith got an A.
 - ψ = #Last semester, all/the foreign students of prof. Smith got an A.

The principle in (i) would predict that ψ should sound better than φ , given that φ has quantification over the entire set of students while the same point could have been made by quantifying on any proper subset; this prediction is not borne out.

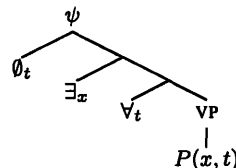
²²The reasoning here needs to be more subtle. In principle, one might get ψ in (131b) to be a scalar alternative of φ in (131a) without any reshuffling, by means of the following strategy. Assume that above each node VP there is a position for a tense operator. That position can be filled or left empty. Assume that when it looks like it is left empty, it is really filled by a “null” tense operator \emptyset_t , that does just nothing. Thus, the proper LF corresponding to φ in (131a) is (i).

(i)



Assume furthermore that $\langle \emptyset_t, \forall_t \rangle$ happens to be a Horn-scale. Thus, I might in principle be allowed to replace the the matrix tense operator \forall_t by \emptyset_t and the embedded null tense operator \emptyset_t by \forall_t , thus getting the LF in (ii), that indeed corresponds to ψ in (131b)

(ii)



this problem by *indirectly* obtaining the negation of ψ in the final strengthened meaning $\text{EXH}(\varphi)$ as the conjunction of the negation of a set of alternatives ψ_P obtained by replacing the indefinite with definite descriptions.

- (131) a. $\varphi = \forall_t[\exists_x P(x, t)]$.
 b. $\psi = \exists_x[\forall_t P(x, t)]$.

This *indirect* strategy predicts the fine status of the minimal variant (101a) of (99a), schematized as φ in (132a). The abstract reason is as follows. This indirect strategy is really only able to derive an “alternative” ψ where the existential quantifier \exists_x has the widest possible scope. In the case of (131), there is only one universal quantifier \forall_t , and thus the net effect is to reverse the scope of \exists_x and \forall_y . In the case of (132), there is another universal quantifier \forall_y besides \forall_t and thus the net effect is that the existential operator \exists_x gets wide scope over both universal operators \forall_t and \forall_y , as in ψ in (132b). Because of the fact that the existential quantifier \exists_x also has wide scope over the universal quantifier \forall_y in ψ , then the implicature that $\neg\psi$ does not contradict common knowledge and φ is not ruled out.

- (132) a. $\varphi = \forall_y[\forall_t[\exists_x P(x, y, t)]]$.
 b. $\psi = \exists_x[\forall_y[\forall_t P(x, y, t)]]$.

One might of course have considered a more *direct* strategy for the case in (131) to start with, as follows. We modify the definition (74) of the set of scalar alternatives to allow for reshuffling of operators. Thus, ψ in (131b) counts as a scalar alternative of φ in (131a), contra (108). The account for the oddness of φ in (131a) thus turns out to be identical to the one suggested in the preceding section 2.1, in terms of the single mismatching scalar implicature $\neg\psi$. This direct approach has two advantages over the indirect approach that I have defended here: first, that the problem with relevance discussed above in the Second Remark disappears; second, that the problem with Percus’ cases of the form (128b) discussed above in the Third Remark disappears too. Yet, this direct approach runs into troubles for the case of (132). In fact, once we have loosened up the definition of the set of scalar alternatives this way by allowing for reshuffling of operators, then we cannot block ψ' in (133b) from counting as an alternative of φ in (132a), repeated in (133a). Contrary to the case of the alternative ψ in (132b), in the case of the alternative ψ' in (133b) the existential quantifier \exists_x has been reshuffled only w.r.t. the universal quantifier \forall_t and not w.r.t. the universal quantifier \forall_y . But then we would get the scalar implicature $\neg\psi'$, namely that “there exists a competition such that there exists no fireman that won it on every day”, which contradicts the common knowledge in (101).

- (133) a. $\varphi = \forall_y \forall_t [\exists_x P(x, y, t)]$.
 b. $\psi' = \forall_y [\exists_x [\forall_t [P(x, y, t)]]]$.

The contrast in (101) is very important for my application to i-predicates in chapter 3, since I will argue that Fox’s effect (8) is formally analogous to the contrast in (101). For this reason, I dismiss this latter direct approach and the idea that the definition (74) of the set of scalar alternatives should be relaxed to allow for reshuffling of operators.

2.3 Extension to downward entailing environments

Let φ and ψ be two contextually equivalent scalar alternatives such that ψ logically asymmetrically entails φ . The proposal presented in the previous section 2.1 predicts φ to sound odd and ψ to sound fine. In other words, if you have to choose between two contextually equivalent alternatives, you

I think that this derivation can be ruled out in one of two ways. One way is that, if the two replacements are done one at the time, then there has got to be an intermediate step at which we are entertaining an alternative such that either its tense argument is unbounded (if we first replace the matrix tense operator \forall_t with the null one \emptyset_t) or else is bound by two different operators (if we first replace the embedded null tense operator \emptyset_t with \forall_t); in either case, the derivation of the alternative ψ from the preadjacent φ contradicts, if not the letter, at least the spirit of the new definition (124) of the set of scalar alternatives. Another way to rule out the derivation just sketched is to deny that there exists any such thing as a Horn-scale of the form $\langle \emptyset_t, \forall_t \rangle$. One might generalize this hypothesis to the claim that no operator is a Horn-mate of the corresponding null/vacuous variant, pace Amsili and Beyssade (2006); see also footnote 25.

should pick the logically stronger one. One would then expect that oddness should flip in downward entailing (henceforth: DE) contexts, namely one would expect the pattern in (134), where O_{DE} is a DE operator.

- (134) ψ is a logically stronger but contextually equivalent scalar alternative to φ .
- a. $O_{DE}(\varphi)$ should sound fine;
 - b. $O_{DE}(\psi)$ should sound odd.

In this section, I test the prediction (134). I note that the data seem to split into two classes w.r.t. prediction (134). And I offer a tentative characterization of the two classes of data, based on the way contextual equivalence is achieved. Throughout this section, I concentrate mainly on the DE context provided by the restrictor of universal quantifiers, since this is the case of interest for my application to i-predicates in chapter 3.

First set of examples Following for instance Sauerland (2004a), let me assume that the masculine gender feature is semantically vacuous. Thus, the universal quantifier in (135a) has a larger restrictor (namely the entire set of Italians) while the universal quantifier in (135b) has a smaller restrictor (namely the subset of Italian women). Thus, sentence (135a) is logically stronger than sentence (135b), despite the fact that they are equivalent given the piece of common knowledge that all Italians come from the same country. And indeed it is the logically weaker sentence (135b) that sounds odd, while the logically stronger sentence (135a) sounds fine.

- (135) a. Gli italiani vengono da un paese bellissimo.
The Italians-MASC come from a country beautiful
'Italians come from a beautiful country'
- b. #Le italiane vengono da un paese bellissimo.
The Italians-FEM come from a country beautiful
'Italian women come from a beautiful country'

The contrast in (136) makes the same point, with a case where the relevant piece of common knowledge is provided by the preceding discourse. The universal quantifier in (136a) has a larger restrictor (namely the set of professors who assigned an A to at least *some* students) and the universal quantifier in (136b) has a smaller restrictor (namely the set of professors that assigned an A to *all* students). Thus, sentence (136a) is logically stronger than sentence (136b), despite the fact that they are equivalent given the piece of common knowledge that, if some professors got a pay raise, then all professors did. And indeed it is the logically weaker sentence (136b) that sounds odd, while the logically stronger sentence (136a) sounds fine.

- (136) Every year, the dean has to decide: if the college has made enough profit that year, he gives a pay raise to every professor who has assigned an A to at least some of his students; if there is not enough money, then no one gets a pay raise.
- a. ... This year, every professor who assigned an A to *some* of his students got a pay raise.
 - b. ... #This year, every professor who assigned an A to *all* of his students got a pay raise.

In conclusion, these contrasts support the prediction (134): out of two contextually equivalent alternatives that differ for a scalar item embedded in a DE contexts, the logically stronger sentence sounds fine while the logically weaker sentence sounds odd.

Second set of examples The universal quantifier in (137a) has a larger restrictor (namely the set of professors who assigned an A to at least *some* of the students) while the universal quantifier in (137b) has a smaller restrictor (namely the set of professors who assigned an A to *all* students). Thus, sentence (137a) is logically stronger than sentence (137b), despite the fact that they are equivalent given the piece of common knowledge that every professor assigns the same grade to all of his students. Surprisingly, in this case it is the logically stronger sentence (137a) that sounds odd, while the logically weaker sentence (137b) sounds fine.

- (137) *Context*: In this department, every professor assigns the same grade to all of his students.
- #This year, every professor of this department who assigned an A to *some* of his students got a prize from the dean.
 - This year, every professor of this department who assigned an A to *all* of his students got a prize from the dean.

The two cases in (138) and (139) make the same point. Again, the logically stronger sentences (138a) and (139a) sound odd while the logically weaker sentences (138b) and (139b) sound fine.²³

- (138) a. #Every father some of whose children have a foreign last name must pay a fine.
 b. Every father whose children have a foreign last name must pay a fine.
- (139) a. #Every student with a blue eye is German.
 b. Every student with blue eyes is German.

In conclusion, these pairs of sentences contradict the prediction (134) that the logically stronger sentence sounds fine while the logically weaker sentence sounds odd.

A generalization In (135)-(139), I have considered pairs of sentences that can be schematized as in (140), where *Strong* is a restrictor of a universal quantifier containing a *strong* scalar item (e.g.: 'all', feminine morphology, etc.) and *Weak* is that same restrictor with the strong scalar item replaced by the corresponding *weak* one (e.g.: 'some', vacuous masculine morphology, etc.).

- (140) a. $\psi = \text{every}(\text{Weak})$
 b. $\varphi = \text{every}(\text{Strong})$.

Of course, ψ logically asymmetrically entails φ . Furthermore the two sentences φ and ψ are contextually equivalent in all the examples considered. Why does φ sound odd and ψ sound fine in the cases (135)-(136) while the opposite happens in the cases (137)-(139)? I want to suggest that the crucial difference hinges on how contextual equivalence between the two alternatives φ and ψ is ensured. To illustrate the idea, consider the minimal case provided by the two pairs (136) and (137). In both cases, the weak and strong restrictors of 'every' are as in (141).

- (141) a. *Weak* = set of professors who assigned an A to *some* students.
 b. *Strong* = set of professors who assigned an A to *all* students.

In the problematic case (137), context entails that the two restrictors *Strong* and *Weak* in (141) are identical in every world w compatible with common knowledge, from which it follows of course that the two sentences φ and ψ are equivalent given common knowledge. In the non-problematic case (136), context ensures that φ and ψ are equivalent without the two restrictors *Strong* and *Weak* in (141) being equivalent too. This observation extends to the other examples considered. In other words: what's special about the problematic cases (137)-(139) is that contextual equivalence between the two alternatives φ and ψ is ensured in a "cheap" way, namely through the contextual equivalence of the two restrictors. I thus suggest the tentative generalization (142).

- (142)
- | | | |
|---|--|----------|
| | $\varphi \leftrightarrow_{w_{ck}} \psi$ | |
| | $\text{Strong} \leftrightarrow_{w_{ck}} \text{Weak} \qquad \text{Strong} \not\leftrightarrow_{w_{ck}} \text{Weak}$ | |
| $\varphi = \text{every}(\text{Strong})$ | (a) fine | (c) odd |
| $\psi = \text{every}(\text{Weak})$ | (b) odd | (d) fine |

In Section 3.2, I'll suggest that the oddness of sentence (5b), with the temporal modifier 'after dinner' and the i-predicate 'tall', provides further evidence for the generalization in the right column of (142); and that the life-time effect triggered by sentence (6b), with the i-predicate 'tall' in the past tense, provides further evidence for the generalization in the left column of (142).

²³Example (139) bears on the issue only under the assumption that the restrictor 'x has blue eyes' means "x has two or more blue eyes" and is thus stronger than 'x has a blue eye'. I am aware that this assumption has been questioned in the recent literature, for instance by Sauerland (2003a) and Spector (2007a). This example might thus be irrelevant.

Toward an account: the idea Assume that the generalization (142) is empirically correct. Is there any way to make sense of it? This is a tough question. Of course, the tough cases are (142a) and (142b) in the left column, namely the cases that do not fit the expected generalization (134). Let me make explicit the nature of the challenge. One might have expected case (142a), with *Strong* in the restrictor of the universal quantifier, to correspond to the strengthened meaning (143a), that entails the negation of the corresponding alternative with embedded *Weak*; but this strengthened meaning would incorrectly predict case (142a) to sound odd, since this strengthened meaning is a contradiction w.r.t. any common knowledge that renders the two restrictors *Strong* and *Weak* contextually equivalent. Furthermore, one might have expected case (142b), with *Weak* in the restrictor of the universal quantifier, to correspond to the strengthened meaning (143b), that has no implicature and thus coincides with the unstrengthened meaning; but this vacuous strengthened meaning would incorrectly predict case (142b) to sound fine.

- (143) a. Naïve strengthened meanings for case (142a):
 $\text{every}(\textit{Strong}) \wedge \neg \text{every}(\textit{Weak})$.
 b. Naïve strengthened meanings for case (142b):
 $\text{every}(\textit{Weak})$.

I want to suggest that both these difficulties disappear once we postulate an exhaustivity operator embedded in the restrictor of the universal operator, besides the matrix one. Let me sketch the idea informally, before I dive into the details. Consider first the case of (142b), that I want to analyze as in (144b). Assume that this sentence contains an exhaustivity operator embedded in the restrictor of the universal quantifier, besides the matrix exhaustivity operator. Assume furthermore that this embedded exhaustivity operator triggers a mandatory implicature $\neg \textit{Strong}$, since the alternative *Strong* is contextually equivalent to the embedded prejacent *Weak*. Then, (142b) is predicted to be bad because it entails ‘ $\text{every}(\textit{Weak} \wedge \neg \textit{Strong})$ ’, that is bad because $\textit{Weak} \wedge \neg \textit{Strong}$ is empty in every world compatible with common knowledge and thus the existence presupposition of ‘every’ cannot be satisfied. Consider next the case of (142a). Assume of course that also this sentence contains an exhaustivity operator embedded in the restrictor of the universal quantifier, besides the matrix exhaustivity operator. The embedded exhaustivity operator does nothing in this case, since its prejacent has no stronger alternatives. Furthermore, the matrix exhaustivity operator does nothing either. In fact, the only matrix alternative that we need to consider is the one obtained by replacing *Strong* by *Weak*, as in (144a). Of course, this alternative retains the embedded exhaustivity operator and thus suffers from presupposition failure, just as the case (144b). Thus, this matrix alternative is in no way contextually equivalent to the matrix prejacent and is therefore not mandatory. Since both matrix and embedded strengthening are vacuous in this case (142a), then this case is correctly predicted to be fine.

- (144) a. (142a) = [EXH [every [EXH *Strong*]]]
 = [$\underbrace{\text{every [EXH *Strong*]}}_{\text{matrix prejacent}}$] and not [$\underbrace{\text{every [EXH *Weak*]}}_{\text{matrix alternative}}$]
 = [every [*Strong*]] and not [every [$\textit{Weak} \wedge \neg \textit{Strong}$]]
 b. (142b) = [EXH [every [EXH *Weak*]]]
 = [EXH [$\underbrace{\text{every [$\textit{Weak} \wedge \neg \textit{Strong}$]}}_{\text{presupposition failure}}$]]

In conclusion, the intuitive idea can be summarized as follows: the case in (142b) sounds odd because the embedded exhaustivity operator screws up the sentence; the case in (142a) sounds fine because the embedded exhaustivity operator screws up the potentially dangerous alternative of the sentence. I am now ready to turn to the details. My proposal hinges on ideas developed in Fox and Spector (2008).

Toward an account: the details In section 2.1, I have concentrated on the matrix case. I have assumed that an instance of the exhaustivity operator is mandatorily adjoined to every matrix sentence, as stated in (86) repeated below in (145). In order to make sense of the hypothesis (145) that

the exhaustivity operator is mandatory in matrix sentences, in section 2.1 I had to partially revise the initial definition of the exhaustivity operator, by letting it depend on a relevance predicate assigned by the context, as in (87b)-(87b') repeated in (146). This new definition (146) of the exhaustivity operator soothes the very strong assumption (145) that the exhaustivity operator is mandatory in matrix clauses: instances of matrix scalar items that trigger no implicature do not contradict the hypothesis (145) that the exhaustivity operator is present, because lack of implicatures can now be attributed to the fact that the corresponding alternatives are not relevant.

(145) The exhaustivity operator EXH is mandatory at matrix scope.

$$(146) \text{ EXH}_{\mathcal{R}}(\varphi) = \varphi \wedge \bigwedge_{\psi \in \mathcal{R} \cap \text{Excl}(\varphi)} \neg\psi$$

I now want to extend assumption (145) from matrix sentences to any scope site, as stated in (147). To make sense of the more general assumption (147), I need to further extend the manoeuvre in (146), as follows. The exhaustivity operator is in essence a universal operator. All (universal) operators come with a restrictor, that is partially determined by grammar and partially by context. The exhaustivity operator is no exception. Thus, let me denote by Ω the restrictor of the exhaustivity operator²⁴ and let me replace the initial definition (73) and the revised definition (146) with the further variant (148). If an LF contains multiple instances of the exhaustivity operator, each instance will come with its own restrictor Ω . For this reason, I will occasionally resort to the more precise notation $\Omega(\varphi)$ that makes explicit the dependence of the restrictor Ω on the prejacent φ . Again, this new definition (148) of the exhaustivity operator soothes the very strong assumption (147), since lack of the exhaustivity operator is equivalent to an exhaustivity operator with an empty restrictor Ω .

(147) The exhaustivity operator EXH is mandatory at every scope site (say, any node that denotes a proposition).

$$(148) \text{ EXH}_{\Omega}(\varphi) = \varphi \wedge \bigwedge_{\psi \in \Omega} \neg\psi$$

Once the two assumptions (147) and (148) are in place, the theory of scalar implicatures can be construed as the theory of the restrictor Ω of the exhaustivity operator. This theory should consist of constraints of two types: constraints on what must belong to Ω and constraints on what cannot belong to Ω . In section 2.1, I have considered examples of both types of constraints. Constraint (149a) says that $\Omega(\varphi)$ cannot contain any alternative that does not belong to the formally defined set $\text{Excl}(\varphi)$ of alternatives excludable w.r.t. φ . Constraint (149b) says that $\Omega(\varphi)$ must contain every excludable alternative that is relevant. Since in section 2.1 I was only concerned with the matrix case, I did not need any other constraint on the domain Ω of the exhaustivity operator besides the two in (149). Thus, I could define Ω as the set $\text{Excl}(\varphi) \cap \mathcal{R}$ of relevant formally excludable alternatives, so that (148) reduces indeed to (146). Once I extend the initial assumption (145) that the exhaustivity operator is present in matrix clauses to the assumption (147) that it is present at every scope site, various complications arise; let me discuss two of them.

- (149) a. $\Omega(\varphi) \subseteq \text{Excl}(\varphi)$;
 b. $\text{Excl}(\varphi) \cap \mathcal{R} \subseteq \Omega(\varphi)$;

Here is a *first* complication. As noted above, I want to posit an exhaustivity operator embedded in the restrictor of a universal operator. To make this assumption compatible with the current statement of assumption (147), that characterizes scope sites as constituents that denote propositions, I will thus resort to a DRT-ish representation, whereby the restrictor of a universal quantifier denotes an open proposition such as *Weak*(x) or *Strong*(x), and the universal operator binds its free variable x . This move also allows me to stick to the assumption that \mathcal{R} is a set of propositions. In section 2.1, I have elaborated a bit on the constraint (149b): I have stated it more explicitly as in (149b') and I have suggested that the latter formulation might actually follow from general plausible assumptions on the relevance property \mathcal{R} , such as those listed in (89). Let me now stipulate without further

²⁴Fox and Spector (2008) use the notation C for the restrictor of the exhaustivity operator, instead of Ω . I cannot use C , because in chapter 3 I will reserve that symbol for the restrictor of Q-adverbs.

discussion that (149b') holds also for the case where the prejacent φ is an embedded constituent such as *Weak(x)* or *Strong(x)*, that denotes a proposition that depends on the assignment of the value to a free variable x .

- (149) b' If the alternative $\psi \in \mathcal{E}xcl(\varphi)$ and the prejacent φ are contextually equivalent, then $\psi \in \Omega(\varphi)$.

Here is a *second* complication. Consider the case of sentence (150a), with a weak scalar item embedded under negation. By assumption (147), this sentence has the LF in (150b), with two exhaustivity operators, one below and one above negation. If the restrictor Ω of the embedded exhaustivity operator were to contain the alternative AND, we would plausibly derive the wrong meaning. Thus, we need some further constraint on the domain Ω of the embedded exhaustivity operator.

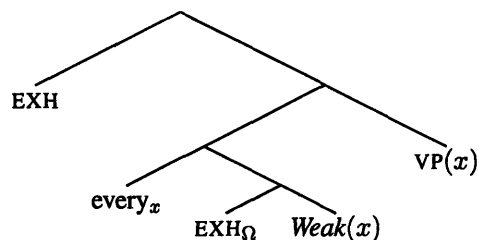
- (150) a. John did't talk to Mary or Sue.
b. [EXH [not [EXH $_{\Omega}$ OR]]]

The somewhat obvious idea is that the domain Ω of the embedded exhaustivity operator in (150b) cannot contain the alternative OR because that instance of the exhaustivity operator occurs in a DE environment, and thus local strengthening leads to global weakening. Fox and Spector (2008) formalize this intuition as a constraint on the domain of the exhaustivity operators. I provide a simplified version of their constraint in (151c), that together with the two constraints repeated in (151a) and (151b) completes my assumptions on the domain Ω of the exhaustivity operator.

- (151) a. $\Omega(\varphi) \subseteq \mathcal{E}xcl(\varphi)$;
b. If the alternative $\psi \in \mathcal{E}xcl(\varphi)$ and the prejacent φ are contextually equivalent, then $\psi \in \Omega(\varphi)$.
c. An LF is ungrammatical if it contains an instance of the exhaustivity operator EXH $_{\Omega}$ with a domain Ω such that there exists a proper subset Ω' of Ω such that the LF obtained by replacing Ω by Ω' is equivalent or stronger.

I now would like to argue that the pattern described in the generalization (142) can be accounted for by the three assumptions (147), (148) and (151). Let me start with the case (142b), namely with the fact that the sentence 'every(*Weak*)' sounds odd in a context that ensures that the two restrictors *Weak* and *Strong* are contextually equivalent. By assumption (147), the LF of this sentence is (152), with a matrix and an embedded exhaustivity operator. Let's concentrate on the embedded exhaustivity operator, whose domain is Ω . By assumption (151a), *Strong* can belong to the domain Ω , since *Strong* belongs to the set of excludable alternatives of the prejacent *Weak*. By assumption (151b), *Strong* must belong to the domain Ω , since we are considering the case where common knowledge renders *Strong* and *Weak* contextually equivalent. The oddness of this case (142b) can then be derived in one of two ways. One way is to attribute it to a contradiction between the two constraints (151b) and (151c): the former constraint wants *Strong* to belong to Ω while the latter prevents *Strong* from belonging to Ω . An alternative way is to let constraint (151b) win over constraint (151c), so that the local implicature is indeed derived and oddness can then be attributed to the fact that the restrictor of 'every' is empty in every world compatible with common knowledge.

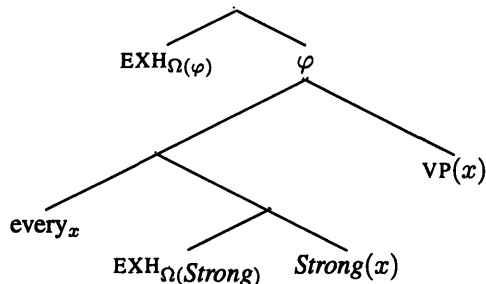
(152)



Let me now turn to the case (142a), whereby the universal quantifier is restricted by *Strong*. Again by assumption (147), the LF corresponding to this case is (153a), with both a matrix and an embedded exhaustivity operator. The set of excludable alternatives of *Strong* is empty; by (151a), the domain $\Omega(\textit{Strong})$ of the embedded exhaustivity operator is thus empty too. The embedded exhaustivity

operator can thus effectively be ignored in computing the truth conditions of the prejacent φ , that therefore turn out to be (153b).

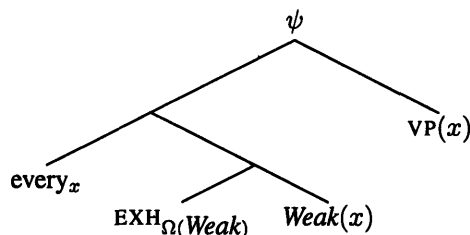
(153) a.



b. $\varphi = \forall_x [Strong(x)][VP(x)]$

By assumption (74) that scalar alternatives are obtained by replacing scalar items with Horn-mates, the prejacent φ of the matrix exhaustivity operator in (153a) has only the scalar alternative ψ in (154a), obtained by replacing *Strong* with *Weak*. By assumption (151b), the alternative *Strong* must belong to the domain $\Omega(Weak)$ of the exhaustivity operator embedded in this alternative ψ , since *Strong* is logically excludable and furthermore by hypothesis contextually equivalent to the prejacent *Weak*. The restrictor of ‘every’ in the alternative ψ thus ends up being empty in every world compatible with common knowledge, namely ψ is an instance of presupposition failure or else a contextual tautology. In either case, ψ is of course not contextually equivalent to the prejacent φ and thus assumption (151b) is moot with respect to the matrix exhaustivity operator in (153a), that thus ends up being vacuous. In conclusion, the truth conditions of the LF (153a) coincide with the truth conditions (153b) of the prejacent φ and the case (142a) is thus correctly predicted to sound fine.²⁵

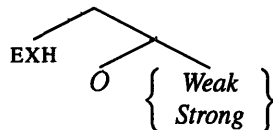
(154) a.



b. $\psi = \forall_x (Weak(x) \wedge \neg Strong(x))$

The cases (142c) and (142d) are fully compatible with the proposal just sketched. Again, these cases will correspond to the two LFs (153a) and (152) respectively. The crucial difference though is that in these two cases the embedded constituents *Weak* and *Strong* are *not* contextually equivalent. Assumption (151b) thus has no bite for the embedded instance of the exhaustivity operator. Thus, I can effectively ignore the embedded exhaustivity operator and consider the simplified LF (155) for all intents and purposes. The predicted strengthened meaning thus are those in (143) and the observed judgments follow straightforwardly into place.

(155)



Extensions So far, I have concentrated on the DE context provided by the restrictor of a universal quantifier. I think that the generalization (142) applies to embedding in any DE environment, not only to embedding in the restrictor of universal quantifiers. For instance, the contrasts (135)-(139)

²⁵ Note that, in order for the account to go through, it is crucial that the alternative ψ in (154) retains the embedded exhaustivity operator. This can be ensured in one of two ways. One way is to assume that there is no Horn-scale that pairs up an operator with a vacuous operator, as already suggested in footnote 22. Another way is to construe assumption (147) as a well-formedness condition that applies to alternatives too. Unfortunately, this assumption that scalar alternatives must retain the embedded exhaustivity operators of the prejacent is at odds with what proposed in Fox and Spector (2008).

remain unaffected if ‘every’ is replaced by ‘no’. To illustrate, I give in (156) and (157) the two cases corresponding to (136) and (137) with ‘every’ replaced by ‘no’, and furthermore with ‘some’ replaced by disjunction in order to circumvent the problem of the PPIs of ‘some’.

- (156) Every year, the dean has to decide: if the college has made enough profit that year, he gives a pay raise to every professor who has taught a graduate or an undergraduate class; if there is not enough money, then no one gets a pay raise.
- This year, no professor who taught a graduate *or* an undergraduate class got a pay raise.
 - #This year, no professor who taught a graduate *and* an undergraduate class got a pay raise.
- (157) *Context: In this department, every professor teaches both a graduate and an undergraduate class in the same field of linguistics.*
- #This year, no professor who taught graduate *or* undergraduate Semantics got a pay raise
 - This year, no professor who taught graduate and undergraduate Semantics got a pay raise.

I thus just rewrite the generalization (142) as in (158), where I have replaced the restrictor of a universal quantifier with an arbitrary DE operator O_{DE} . Before I move on, let me make a couple of remarks on the special case where O_{DE} is negation.

(158)

	$\varphi \leftrightarrow_{w_{ck}} \psi$	
	$Strong \leftrightarrow_{w_{ck}} Weak$	$Strong \not\leftrightarrow_{w_{ck}} Weak$
$\varphi = O_{DE}(Strong)$	(a) fine	(c) odd
$\psi = O_{DE}(Weak)$	(b) odd	(d) fine

A few authors have observed that odd sentences remain odd when embedded under negation. For instance Spector (2007a) construes the plurality inference triggered by plural morphology as a scalar implicature. The oddness of sentence (159a) is thus due to the fact that this implicature mismatches with the piece of common knowledge that people can marry only one person at the time. Spector notes that embedding under negation does not affect the oddness of (159a), as shown in (159b). But the generalization (158) says that negation is not the interesting case to test the behavior of oddness in DE contexts. In fact, negation has the property that the right column of the table (158) can never hold, since the hypothesis that the matrix sentences $\varphi = \neg Strong$ and $\psi = \neg Weak$ are contextually equivalent immediately entails that the corresponding embedded sentences *Strong* and *Weak* are contextually equivalent too.

- (159) a. #Last summer, Mario married (some) Italian girls.
b. #Despite his family’s pressure, Mario didn’t marry Italian girls.

Further support for the generalization (158) comes from the contrast in (160).²⁶ Sentence (160a) contains the strong scalar item ‘all’ embedded under negation. This case falls of course under the left column of table (158), since context ensures the equivalence between the embedded sentence *Strong* = ‘all professors assigned an A’ and the alternative *Weak* = ‘some professors assigned an A’. Thus, the acceptability of sentence (160a) fits with the generalization (158). Sentence (160b) is truth-conditionally equivalent to sentence (160a), and yet sounds odd. This is expected if ‘not all’ is a single scalar item: in this case, table (158) does not apply, because there is no embedding in a DE environment here; hence, sentence (160b) is predicted to sound odd, since it is contextually equivalent to but logically weaker than the alternative with ‘not all’ replaced by ‘no’.

- (160) In this department, all professors get together at the end of the semester and decide a grade to assign to all of their students.
- It is false that [this year all professors assigned an A].

²⁶This example came up in conversation with E. Chemla.

b. ?This year, not all professors assigned an A.

So far, I have considered the case of embedding under a DE operator O_{DE} . I think that the generalization (158) can be extended to upward entailing (henceforth: UE) operators too, as stated in the revised generalization (161).²⁷

$$(161) \quad \begin{array}{c} \varphi \leftrightarrow_{wck} \psi \\ \overbrace{\text{Strong} \leftrightarrow_{wck} \text{Weak}} \quad \overbrace{\text{Strong} \not\leftrightarrow_{wck} \text{Weak}} \\ \underbrace{\hspace{15em}} \\ \begin{array}{ccc} O \text{ is DE} & & O \text{ is UE} \end{array} \\ \begin{array}{l} \varphi = O(\text{Strong}) \\ \psi = O(\text{Weak}) \end{array} \quad \begin{array}{|c|c|c|} \hline \begin{array}{l} \text{(a) fine} \\ \text{(b) odd} \end{array} & \begin{array}{l} \text{(c) odd} \\ \text{(d) fine} \end{array} & \begin{array}{l} \text{(e) fine} \\ \text{(f) odd} \end{array} \\ \hline \end{array} \end{array}$$

The idea behind this generalization (161) can be informally spelled out as follows. We have to choose between these two alternatives φ and ψ , obtained by embedding under an operator O either *Strong* or *Weak*. We start from the embedded context and move up. If the two embedded constituents are *indeed* contextually equivalent, then the choice is made at the embedded level: we pick the sentence with the logically stronger embedded constituent *Strong*, no matter the monotonicity of the embedding operator O . If the two embedded constituents are *not* contextually equivalent, then we cannot make the choice at the embedded level and need instead to look one level up. In this case, the monotonicity of the embedding operator O does of course matter. In fact, we pick the alternative that is globally logically stronger. In other words, I assume that also in the case of a UE embedding operator O , the decision is made at the embedded level whenever the embedded constituents are contextually equivalent. Of course, in the case where O is a UE operator, it does not make any difference whether the choice is made locally or globally, because the two options lead to the same conclusion, since the globally logically stronger alternative is the one with the logically stronger embedded alternative.

Remark: back to the cases considered in sections 2.1 and 2.2 As noted above, (147)-(148) reduce to (145)-(146) in the case of matrix sentences. Thus, the modifications suggested in this section have no effect for the basic matrix case considered in section 2.1. Yet, they do have consequences for the case of the two sentences (101a) and (101b) considered in section 2.2. The case of sentence (101a) remains unchanged, as shown in detail in (162). By assumption (147), this sentence now receives the LF (162a), with a matrix and an embedded exhaustivity operator.²⁸ Here, I am using \forall_t as a shorthand for ‘every day’, \exists_x as a shorthand for ‘a fireman’ and $P(x, t)$ as a shorthand for ‘ x won the running competition on day t ’. I will stick to my crucial assumption that ⟨‘a fireman’, ‘the fireman P ’⟩ (now abbreviated as $\langle \exists_x, x \rangle$) is a Horn-scale. Of course, also ⟨‘a fireman’, ‘every fireman’⟩ is a Horn-scale; but the latter scale is irrelevant to my reasoning, and I will thus ignore it. The truth conditions of the constituent $B(t)$ are (162b): none of the alternatives obtained by replacing \exists_x by x is excludable and the embedded exhaustivity operator is therefore vacuous. Let $B_{[\exists_x/x]}(t)$ be the constituent $B(t)$ with the existential quantifier \exists_x replaced by x . The truth conditions of $B_{[\exists_x/x]}$ are (162c).²⁹ The truth conditions of the matrix constituent A can then be computed as in (162d): in step (i), I have used the fact that the only non-maximal scalar item in the scope of the matrix exhaustivity operator is the existential quantifier \exists_x , that participates in the

²⁷Thanks to Emmanuel Chemla (p.c.) for discussion on this point.

²⁸Strictly speaking, assumption (147) predicts a third exhaustivity operator in between the existential quantifier \exists_x and the atomic formula $P(x, t)$. But this operator would be vacuous (because there would be no scalar items in its scope), and I thus ignore it.

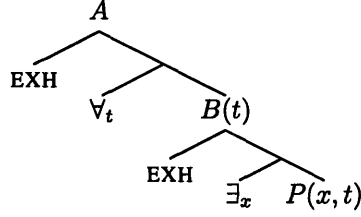
²⁹I am possibly making an over simplification in (162c). Perhaps, ⟨‘the fireman P ’, ‘the fireman P' ’⟩ is a Horn-scale for any two properties P and P' (here abbreviated $\langle x, x' \rangle$). Under this assumption, (162c) and (162d) become (162c') and (162d') below.

$$(162) \quad \text{c}' \quad B_{[\exists_x/x]}(t) = P(x, t) \wedge \bigwedge_{x' \neq x} \neg P(x', t) \stackrel{\text{def.}}{=} \text{ONLY}_x P(x, t)$$

Horn-scale $\langle \exists_x, x \rangle$; in step (ii), I have used the results in (162b) and (162c). The result obtained at the end of (162d) is exactly the same one obtained in section 2.2. Thus the embedded exhaustivity operator required by assumption (147), does not affect my previous proposal.

(162) Every day, a fireman won the running competition.

a.



b. $B(t) = \exists_x P(x, t)$

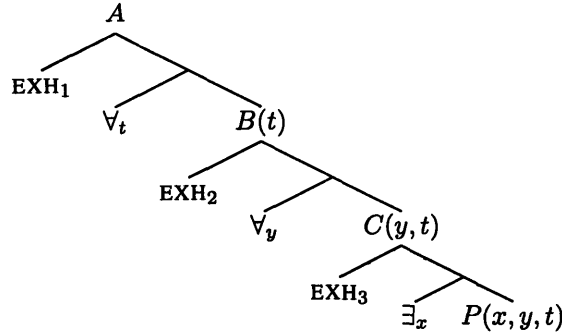
c. $B_{[\exists_x/x]}(t) = P(x, t)$

d. $A \stackrel{(i)}{=} (\forall_t B(t)) \wedge \bigwedge_x \neg (\forall_t B_{[\exists_x/x]}(t))$
 $\stackrel{(ii)}{=} (\forall_t \exists_x P(x, t)) \wedge \bigwedge_x \neg (\forall_t P(x, t))$
 $= \forall_t \exists_x P(x, t) \wedge \neg \exists_x \forall_t P(x, t)$

Let me now turn to the case of sentence (101b), discussed in (163). This case is more complicated because there are three operators: a universal operator over competitions \forall_y , a universal operator over days of the week \forall_t and an existential quantifier over firemen \exists_x . By assumption (147), the corresponding LF thus has three exhaustivity operators, as in (163a). Again as in the preceding case, the embedded exhaustivity operator EXH₃ is vacuous, and thus the denotation of the constituent $C(y, t)$ are (163b). Let $C_{[\exists_x/x]}(y, t)$ be the constituent $C(y, t)$ with the existential quantifier \exists_x replaced by x , whose corresponding truth conditions are (163c). The denotation of the constituent $B(t)$ can then be computed as in (163d): in step (i), I have used the fact that the only non-maximal scalar item in the scope of the matrix exhaustivity operator EXH₂ is the existential quantifier \exists_x , that participates in the Horn-scale $\langle \exists_x, x \rangle$; in step (ii), I have used the results in (163b) and (163c). Let $B_{[\exists_x/x]}(t)$ be the constituent $B(t)$ with the existential quantifier \exists_x replaced by x , whose corresponding truth conditions are (163e). The denotation of the matrix constituent A can then be computed as in (163f): in step (i), I have used the fact that the only non-maximal scalar item in the scope of the matrix exhaustivity operator EXH₁ is the existential quantifier \exists_x , that participates in the Horn-scale $\langle \exists_x, x \rangle$; in step (ii), I have used the results in (163d) and (163e); in step (iii), I have noted that (*) entails (**), so that the latter can be dropped.

(163) Every day, for every competition, a fireman won.

a.



d' $A = (\forall_t B(t)) \wedge \bigwedge_x \neg (\forall_t B_{[\exists_x/x]}(t))$
 $= (\forall_t \exists_x P(x, t)) \wedge \bigwedge_x \neg (\forall_t \text{ONLY}_x P(x, t))$
 $= \forall_t \exists_x P(x, t) \wedge \neg \exists_x \forall_t \text{ONLY}_x P(x, t)$

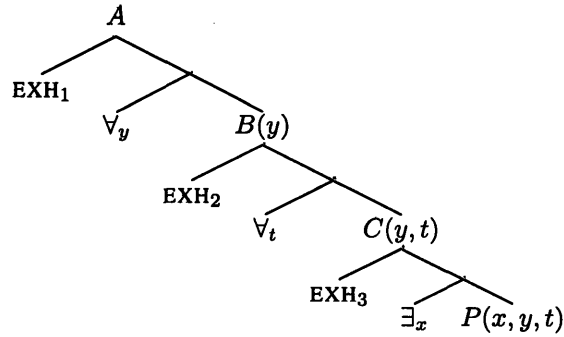
I think that this modification does not affect my proposal.

$$\begin{aligned}
\text{b. } & C(y, t) = \exists x P(x, y, t) \\
\text{c. } & C_{[\exists x/x]}(y, t) = P(x, y, t) \\
\text{d. } & B(t) \stackrel{(i)}{=} (\forall y C(y, t)) \wedge \bigwedge_x \neg (\forall y C_{[\exists x/x]}(y, t)) \\
& \stackrel{(ii)}{=} (\forall y \exists x P(x, y, t)) \wedge \bigwedge_x \neg (\forall y P(x, y, t)) \\
& = (\forall y \exists x P(x, y, t)) \wedge \neg (\exists x \forall y P(x, y, t)) \\
\text{e. } & B_{[\exists x/x]}(t) = \forall y P(x, y, t) \\
\text{f. } & A \stackrel{(i)}{=} \forall_t B(t) \wedge \bigwedge_x \neg (\forall_t B_{[\exists x/x]}(t)) \\
& \stackrel{(ii)}{=} \forall_t [(\forall y \exists x P(x, y, t)) \wedge \neg (\exists x \forall y P(x, y, t))] \wedge \bigwedge_x \neg (\forall_t \forall y P(x, y, t)) \\
& = \forall_t \forall y \exists x P(x, y, t) \wedge \underbrace{\neg \exists_t \exists x \forall y P(x, y, t)}_{(*)} \wedge \underbrace{\neg \exists_x \forall_t \forall y P(x, y, t)}_{(**)} \\
& \stackrel{(iii)}{=} \forall_t \forall y \exists x P(x, y, t) \wedge \neg \exists_t \exists x \forall y P(x, y, t)
\end{aligned}$$

In conclusion, (163f) says that sentence (101b) triggers the scalar implicature $\neg \exists_t \exists x \forall y P(x, y, t)$ that “it is not the case that there is a day on which a single firemen won all of the competitions.” This implicature is different from the scalar implicature $\neg \exists x \forall_t \forall y P(x, y, t)$ that was obtained in section 2.2 for the LF with a single matrix exhaustivity operator. Yet, the implicature obtained here is innocuous, since it is indeed compatible with the common knowledge in (101), thus predicting the felicity of the sentence (101b). Yet, things are different for the variant of the sentence (101b) with the opposite relative scope of the two universal quantifiers, corresponding to the LF (164a) with truth conditions (164b). In this case in fact, the implicature $\neg \exists_y \exists x \forall_t P(x, y, t)$ that “there exists no competition that was won by the same fireman on every day” does mismatch with the common knowledge in (101).

(164) For every competition, every day a fireman won.

a.



$$\text{b. } A = \forall_t \forall y \exists x P(x, y, t) \wedge \neg \exists_y \exists x \forall_t P(x, y, t)$$

I can thus conclude as follows. If sentence (101b) admits both LFs (163) and (164), then it is correctly predicted to sound fine in the context considered, because it admits at least one LF, namely (163), which is fine. If instead the surface scope of the two universal quantifiers determines their LF scope, then my proposal predicts that sentence (163) should be fine in the context considered while the variant in (164) with the opposite surface scope of the two universal quantifiers should sound odd.

2.4 Extension to Presuppositions

This section somewhat tentatively explores the hypothesis that anti-presuppositions can be derived as plain scalar implicatures. This hypothesis is developed through various steps.

First step As noted in section 1.2, an important class of inferences that have been studied in the literature is that of ANTI-PRESUPPOSITIONS. These inferences are exemplified in (165)-(166): sentence (165a) triggers the inference that the factivity presupposition of the corresponding sentence (165b) is not satisfied; analogously, sentence (166a) triggers the inference that the uniqueness presupposition of the corresponding sentence (166b) is not satisfied. There seems to be wide agreement that the anti-presuppositions triggered by sentences (165a) and (166a) do indeed arise by “competition” with the corresponding alternatives (165b) and (165b). The crucial issue now is to understand what is the nature of the “competition” that triggers these anti-presuppositions.

- (165) a. John *believes* that Mary is pregnant.
 b. John *knows* that Mary is pregnant.
- (166) a. $\varphi = A$ mistress of the victim arrived late.
 b. $\psi = The$ mistress of the victim arrived late.

As reviewed in section 1.2, it has been suggested in the literature that anti-presuppositions cannot be either scalar implicatures or presuppositions. Thus, a new, dedicated principle has been posited in order to steer the competition that underlies these inferences, such as MP (34). Sauerland (2008) provides the convincing empirical argument (30) to show that anti-presuppositions are not presuppositions, based on their different projection behavior under universal quantifiers. On the other hand, the argument (32) against treating anti-presuppositions as run-of-the-mill scalar implicatures was contingent on the adoption of a pragmatic theory of scalar implicatures. That argument thus evaporates once the pragmatic theory is replaced with a grammatical theory of scalar implicatures such as the one reviewed in section 2.1, whereby scalar implicatures are derived through a covert but syntactically realized operator akin to overt ‘only’. Furthermore, the anti-presuppositions of sentences (165a) and (166a) can be brought out with an overt ‘only’ as in (167), just as it is the case for run-of-the-mill scalar implicatures, as noted in (68)-(70) from Fox (2007a).

- (167) a. John only believes_F that Mary is pregnant.
 b. You are only one_F mistress of John.

I thus want to entertain the hypothesis that anti-presuppositions are not a separate, peculiar class of inferences but rather run-of-the-mill scalar implicatures. Here is an initial way to cash out this idea. Suppose that a sentence φ that contains presuppositional items denotes two propositions, namely its PRESUPPOSITION φ_{prs} and its ASSERTION φ_{asr} . Thus, $\llbracket\varphi\rrbracket = \langle\varphi_{prs}, \varphi_{asr}\rangle$. How are we going to interpret the LF (168) obtained by appending the exhaustivity operator EXH to sentence φ then?

- (168) $\llbracket[\text{EXH } \varphi]\rrbracket = ?$

Here is a very minimal way to go, that I illustrate by concentrating on the pair of sentences φ and ψ in (166). There is wide consensus on what the presupposition ψ_{prs} of sentence ψ in (166b) should look like, namely the set of worlds where the victim had one and only one mistress. But what should the assertion ψ_{asr} look like? It seems to me that both options (169) are in principle available. Option (169b) says that the two sentences in (166) have the same assertion, namely $\psi_{asr} = \varphi_{asr}$. Option (169a) says that the assertion ψ_{asr} is a subset of the presupposition ψ_{prs} . If we were to construe the denotation of ψ as a partial function, then option (169a) would correspond to defining ψ_{asr} as the set of worlds where the function is both defined and true. I think that the two options (169) do *not* make different predictions for projection behavior.

- (169) a. $\psi_{asr} =$ the set of worlds where the victim has a unique mistress and she arrived late.
 b. $\psi_{asr} =$ the set of worlds where the victim has a mistress who arrived late.

The choice between the two variants in (169) has important consequences for the issue of the proper derivation of anti-presuppositions. Suppose that the right definition of the assertion ψ_{asr} of sentence ψ in (166b) were (169a). And assume that the proper solution to problem (168) is (170). This is the minimal modification that we can make to the current framework in order to make it compatible with the assumption that the prejacent φ denotes not one but two propositions. This modification says that the exhaustivity operator ignores the presupposition component and only works at the level

of assertion. The anti-presupposition of sentence φ in (166a) follows straightforwardly from the assumption that \langle 'a', 'the' \rangle is a Horn-scale.

$$(170) \quad \llbracket \text{EXH } \varphi \rrbracket = \langle \varphi_{\text{pres}}, \text{EXH}(\varphi_{\text{asr}}) \rangle.$$

Let me take stock. An anti-presupposition looks suspiciously similar to a scalar implicature, with the special property that the relevant alternative contains a presuppositional item. Within the pragmatic framework to scalar implicatures, anti-presuppositions cannot be derived as scalar implicatures, because of the issue in (32). Thus, basically the same engine needs to be used twice: in the form of a pragmatic derivation of scalar implicatures; and in the form of a grammatical derivation of anti-presuppositions. Once we grammaticalize scalar implicatures to start with, there is no reason any more to treat the two types of inferences separately.

Second step The odd sentences in (60)-(64) containing 'some' remain odd if 'some' is replaced by 'every'/'all', as shown in (171)-(175). Example (172) is due to Percus (2001); example (173) is due to Spector (2007b).

- (171) Mary is conducting a survey on names and last names of Italian children. She knows that all children inherit the last name of their father; hence, all children of a given couple share the same last name. This week, she has interviewed the children of five couples, A through E, in order to record their names and last names...
- a. φ = #Every child of couple C has a French last name.
 - b. ψ = The children of couple C have a French last name.
- (172) Alluding to the Marx Brothers:
- a. #Each one's mother was named Minnie.
 - b. Their mother was named Minnie.
- (173) Speaking of the atoms of a molecule which cannot be separated:
- a. #Every atom went right.
 - b. The atoms went right.
- (174) #Both eyes of the victim are blue.
- (175) a. #Every Italian comes from a beautiful country.
b. Italians come from a beautiful country.

The account proposed in section 2.1 for the oddness of the original sentences with 'some' of course does not extend to these variants (171a)-(175a), since these latter sentences plausibly express the strongest meaning among their alternatives and thus (blind) strengthening is vacuous. I want to suggest that the oddness of sentences (171a)-(175a) is due to the fact that, nonetheless, they do not carry the strongest presupposition. Let me illustrate the idea, by concentrating for example on the pair of sentences φ and ψ in (171). Let me assume that the plural definite article and universal quantifiers are Horn-mates, as stated in (176).³⁰ Thus, let me consider the alternative ψ in (171b), where 'every' is replaced by 'the'.

- (176) \langle 'every', 'the' \rangle is a Horn-scale

Let me pause to discuss the proper semantics of this alternative ψ . Distributive predication with definite subjects, as in the case of ψ , requires the predicate to be operated upon by the DISTRIBUTIVE OPERATOR Dist . Following for instance Link (1983), I take the function $\text{Dist}(\llbracket \text{VP} \rrbracket^w)$ to be true of a plurality x iff the property $\llbracket \text{VP} \rrbracket^w$ holds of all the atomic parts y of x (i.e. $y \leq_{\text{AT}} x$). Thus, φ and

³⁰I think this assumption is plausible, for the following reason. Every existing account for the oddness of Hawinks' sentences (10) needs the assumption that 'a' and the singular definite article are alternatives. Since Sharvy (1980), it has become standard to assume that there is no difference between singular and plural definites. Thus, we need the assumption that existential quantifiers and definites are Horn-mates. If Horn-mateness is transitive and if furthermore existential and universal quantifiers are Horn-mates, then it follows that universal quantifiers and definites are Horn-mates too, as stated in (176).

ψ end up having the same meaning. Nonetheless, they have different presuppositions. In fact, the distributivity operator **Dist** does something more: it also introduces the so called **HOMOGENEITY PRESUPPOSITION**, namely the presupposition that the property $[[VP]^w$ either holds of all the atomic parts y of x or else it does not hold of any of them; see Fodor (1970), Löbner (1985), von Stechow (1997) and Gajewski (2005), among others. Thus, the proper semantics of the distributive operator **Dist** is that in (177). The homogeneity presupposition can be detected for example by means of negation: the sentence ‘Sue didn’t see the boys’ conveys that Sue didn’t see any of the boys, which differs from the plain meaning (namely that “Sue didn’t see every boy”) but does follow from the plain meaning plus the homogeneity presupposition.

$$(177) \quad \text{Dist}([VP]^w) = \lambda x : YES^w(x) \vee NO^w(x) \cdot YES^w(x),$$

- a. $YES^w(x) = 1$ iff $[[VP]^w(y)$ for every y such that $y \leq_{AT} x$,
- b. $NO^w(x) = 1$ iff $\neg[[VP]^w(y)$ for every y such that $y \leq_{AT} x$.

By (177), ψ in (171b) bears the homogeneity presupposition ψ_{prs} in (178b), according to which either all the children of couple C have a French last name or else none of them does. No such presupposition is carried by φ in (171a), as stated in (178a), where I am ignoring other potential presuppositions which are irrelevant to my point.

$$(178) \quad \begin{array}{l} \text{a. } \varphi_{prs} = \mathcal{W}. \\ \text{b. } \psi_{prs} = YES \cup NO, \\ \quad \text{i. } YES = \{w \mid \text{every child of the couple has a long last name } w\}, \\ \quad \text{ii. } NO = \{w \mid \text{no child of the couple has a long last name in } w\}. \end{array}$$

At this point, I would like to suggest that the oddness of sentence φ in (171a) arises because of the anti-homogeneity inference triggered by competition with ψ in (171b). Unfortunately, there is no way to get this inference by sticking with the simple assumption (170) that strengthening is only performed at the level of assertion. Let me make the problem explicit. The trick used in the case of the pair φ and ψ in (166) was to effectively *define* the assertion of the alternative ψ as the conjunction of the assertion of φ with the presupposition of ψ , as stated in (179). Thus, the conjunction $\varphi_{asr} \wedge \neg\psi_{asr}$ obtained by strengthening the assertion entails $\neg\psi_{prs}$. But this trick based on defining the assertion of the alternative ψ as in (179) does not work in the case of the pair φ and ψ in (171): the assertion φ_{asr} of φ entails the homogeneity presupposition ψ_{prs} of ψ ; thus, ψ_{asr} defined in (179) boils down to φ_{asr} ; hence, no strengthening of the assertion happens in this case.

$$(179) \quad \psi_{asr} = \psi_{prs} \wedge \varphi_{asr}$$

In order to get out of trouble, I would like to suggest that (170) be replaced by (180), namely that strengthening be performed not just at the level of assertion but also at the level of presupposition.³¹

³¹A similar claim has been made recently in Sharvit and Gajewski (2007) and Gajewski and Sharvit (2009). Building on Simons (2006), they note the following fact: sentence (i) triggers the implicature in (ia) but *not* the one in (ib), contrary to other attitude verbs, such as ‘to be certain’. They suggest that this behavior has to do with the fact that ‘some’ is embedded in a UE environment if we consider the presupposition but in a DE environment if we consider the assertion. And they claim that, in order to capture this intuition, it is necessary to assume that strengthening be performed independently at the level of the presupposition, as in (180) — even though they adopt a very different framework.

- (i) $\varphi =$ John is sorry that A or B.
 - a. \rightsquigarrow John believes that A or B and not both
 - b. $\not\rightsquigarrow$ John is sorry that A or B and not both.

But I think that this fact (i) can be obtained also by strengthening just at the level of the assertion as in (170), and thus do not really provide an argument in favor of strengthening also at the level of the presupposition as in (180). Here is why. They assume that the sentence (i) denotes the partial function in (ii), with the standard notation of Heim and Kratzer (1998). Thus, the assertion φ_{asr} can only be defined as in (iii) as the set of worlds where the function φ is defined and equal to 1. Let ψ_{asr} be the assertion of the corresponding alternative obtained by replacing ‘or’ with ‘and’, as in (iiib).

$$(ii) \quad \varphi = \lambda w : DOX(w) \subseteq p \cdot BUL(w) \subseteq \neg p$$

$$(iii) \quad \text{a. } \varphi_{asr} = \left(DOX(w) \subseteq OR \right) \wedge \left(BUL(w) \subseteq \neg OR \right)$$

$$(180) \quad \llbracket [\text{EXH } \varphi] \rrbracket = \langle \text{EXH}_{\text{prs}}(\varphi_{\text{prs}}), \text{EXH}_{\text{asr}}(\varphi_{\text{asr}}) \rangle.$$

The most straightforward way to make (180) explicit would be (181). Let the set of scalar alternatives $\text{Alt}(\varphi)$ be defined as in (74)³² both for the computation of the strengthened presupposition and for the computation of the strengthened meaning, namely as the set of LFs obtained from the prejacent by replacing scalar items with Horn-mates. Let the set $\text{Excl}_{\text{prs}}(\varphi)$ of alternatives presupposition-excludable and the set $\text{Excl}_{\text{asr}}(\varphi)$ of alternatives assertion-excludable be defined as in (75)³³: $\text{Excl}_{\text{prs}}(\varphi)$ is the set of those alternatives whose presupposition asymmetrically entails the presupposition of φ and $\text{Excl}_{\text{asr}}(\varphi)$ is the set of those alternatives whose assertion asymmetrically entails the assertion of φ . Finally, let the strengthened presupposition $\text{EXH}_{\text{prs}}(\varphi)$ and the strengthened assertion $\text{EXH}_{\text{asr}}(\varphi)$ be defined as in (73): $\text{EXH}_{\text{prs}}(\varphi)$ is the presupposition of φ conjoined with the negation of all the excludable presuppositions and $\text{EXH}_{\text{asr}}(\varphi)$ is the assertion of φ conjoined with the negation of all the excludable assertions.

(181) PRESUPPOSITION	STRENGTHENED ASSERTION
$\text{Alt}(\varphi)$	$\text{Alt}(\varphi)$
$\text{Excl}_{\text{prs}}(\varphi) = \{\psi \in \text{Alt}(\varphi) \mid \psi_{\text{prs}} \xrightarrow{a} \varphi_{\text{prs}}\}$	$\text{Excl}_{\text{asr}}(\varphi) = \{\psi \in \text{Alt}(\varphi) \mid \psi_{\text{asr}} \xrightarrow{a} \varphi_{\text{asr}}\}$
$\text{EXH}_{\text{prs}}(\varphi) = \varphi_{\text{prs}} \wedge \bigwedge_{\psi \in \text{Excl}_{\text{prs}}(\varphi)} \neg \psi_{\text{prs}}$	$\text{EXH}_{\text{asr}}(\varphi) = \varphi_{\text{asr}} \wedge \bigwedge_{\psi \in \text{Excl}_{\text{asr}}(\varphi)} \neg \psi_{\text{asr}}$

Assume that also strengthening at the level of presupposition is blind to common knowledge, as stated in (182), which I'll refer to as the BH_{prs} .

(182) The notion of entailment relevant for the computation of the strengthened presupposition $\text{EXH}_{\text{prs}}(\varphi)$ of a sentence φ is that of *logical* entailment rather than that of entailment *relative to common knowledge*.

Once (180)-(182) are in place, the account for the oddness of sentence φ in (171a) follows straightforwardly. Of course, ψ_{prs} in (178b) logically asymmetrically entails φ_{prs} in (178a). Hence, the set of alternatives presupposition-excludable w.r.t. φ is $\text{Excl}_{\text{prs}}(\varphi) = \{\psi\}$ and the blind strengthened presupposition of φ thus boils down to (183).³⁴ This strengthened presupposition contradicts the piece of common knowledge that all the children of a given couple inherit their father's last name and thus share the same last name. Note the crucial role played by the BH_{prs} (182): if the strengthened presupposition were computed using entailment relative to common knowledge, then ψ could not count as a presupposition-excludable alternative w.r.t. φ , since φ_{prs} and ψ_{prs} are equivalent given the common knowledge that all the children of a given couple share the same last name. This line of account trivially extends to the other cases (171)-(175).³⁵ In Section 3, I'll argue that the ban against

$$\text{b. } \psi_{\text{asr}} = \left(\text{DOX}(w) \subseteq \text{AND} \right) \wedge \left(\text{BUL}(w) \subseteq \neg \text{AND} \right)$$

Then, by strengthening only the assertion as in (170), I would derive the strengthened assertion in (iv), where in the last step I have used the fact that $\text{BUL}(w) \subseteq \neg \text{OR}$ entails $\text{BUL}(w) \subseteq \neg \text{AND}$.

$$\begin{aligned} \text{(iv) } \text{EXH}(\varphi_{\text{asr}}) &= \varphi_{\text{asr}} \wedge \neg \psi_{\text{asr}} \\ &= \left(\text{DOX}(w) \subseteq \text{OR} \right) \wedge \left(\text{BUL}(w) \subseteq \neg \text{OR} \right) \wedge \neg \left[\left(\text{DOX}(w) \subseteq \text{AND} \right) \wedge \left(\text{BUL}(w) \subseteq \neg \text{AND} \right) \right] \\ &= \left(\text{DOX}(w) \subseteq \text{OR} \wedge \neg \text{DOX}(w) \subseteq \text{AND} \right) \wedge \text{BUL}(w) \subseteq \neg \text{OR} \end{aligned}$$

By (iv), I get the implicature $\text{DOX}(w) \not\subseteq \text{AND}$ that is slightly weaker than the implicature $\text{DOX}(w) \subseteq \neg \text{AND}$ in (ia), but could be strengthened to (ia) by virtue of a pragmatic reasoning such as the one suggested by Russell (2006). I thus tentatively conclude that the case in (i) can be handled by strengthening only the assertion as in (170), without any need for strengthening the presupposition, as in (180).

³²Or, more properly, as in (124).

³³Or, more properly, as in (106)

³⁴Here, I am ingoring once more the issue of relevance, and thus why the mismatching scalar presupposition is mandatory.

³⁵Let me specifically discuss the case of sentence (174). I think it is very plausible to assume that 'both' has the same semantics as the universal quantifier 'every', only with the extra DUALITY PRESUPPOSITION that its restrictor has cardinality

a universal adverb such as ‘always’ in the case of *i*-predicates, as illustrated in (3b), can be accounted for in the same way as I have accounted here for the oddness of sentence (171a), by exploiting the fact that English has a covert Q-adverb GEN which triggers the homogeneity presupposition too.³⁶

$$(183) \text{ EXH}_{\text{prs}}(\varphi) = \varphi_{\text{prs}} \wedge \neg\psi_{\text{prs}} = \neg\psi_{\text{prs}}.$$

Third step The account that I have in place right now runs into two problems. One problem concerns fine sentences containing a universal quantifier, such as (184a): the current proposal predicts this sentence to presuppose the negation of the homogeneity presupposition of (184b) despite the fact that its assertion entails that presupposition!

- (184) a. Every friend of mine is tall.
b. My friends are tall.

Another problem concerns the following observation due to Sauerland (2003b): assume that ‘a’ and ‘every’ are Horn-mates and that ‘every’ carries an existence presupposition that its restrictor be not empty; then, sentence (185a) would trigger the strengthened presupposition that the existence presupposition of (185b) is false; sentence (185a) would thus contradictorily presuppose that there are no whales and entail that there is one.

- (185) a. Mary saw a whale.
b. Mary saw every whale.

two. Assumption (176) that ‘every’ and ‘the’ are Horn-mates thus plausibly extends to the assumption that ‘both’ and ‘the’ are Horn mates. The account just suggested for the oddness of the other sentences (171a)-(175a) thus straightforwardly extends to the case of sentence (174): its strengthened presupposition entails the negation of the homogeneity presupposition triggered by the corresponding alternative (i) and thus mismatches with the piece of common knowledge that people have eyes of the same color. So far so good. But as soon as we let ‘both’ and ‘the’ be alternatives, we immediately run into the following problem: why is it that sentence (i) does not sound odd? In fact, if ‘both’ and ‘the’ are Horn-mates, then the strengthened presupposition of sentence (i) should entail the negation of the duality presupposition of ‘both’ and thus should mismatch with the common knowledge that people have two eyes.

- (i) The eyes of the victim are blue.

Perhaps, one way to get out of trouble might be to assume that ‘both’ and ‘the’ are not Horn-mates to start with. Rather the proper Horn-mate of ‘both’ is ‘the two’. Under this assumption, the oddness of sentence (174) is still accounted for by repeating the preceding account with the alternative (ii) instead of (i). Furthermore, sentence (i) is now protected by the fact that it has no relevant presuppositionally stronger alternative; and sentence (ii) is protected by the fact that it has the strongest possible presupposition, namely both the duality and the homogeneity presupposition.

- (ii) The two eyes of the victim are blue.

Yet, this assumption that the single lexical item ‘both’ is an alternative of the composite expression ‘the two’ is far from trivial. For one thing, ‘the two’ is plausibly not a constituent. Furthermore, this assumption is not compatible with the idea that scalar alternatives cannot be syntactically “more complex” than the target sentence, as suggested for example by Katzir (2008). I leave the issue open.

³⁶Let me consider the two variants of Chemla’s example (62) in (i) and (ii). In both cases, the (a) sentence with ‘some’ sounds odd. The interesting cases are the (b) and the (c) sentences. The proposal developed so far makes the following prediction: the (b) sentence in (ii) should sound worse than the (b) sentence in (i), since in the case in (ii) the homogeneity presupposition applies while in the case of (i) it does not apply. The same asymmetry is predicted for the case of the (c) sentences.

- (i) [some/most] Because of an MIT policy against intellectual discrimination, every professor has to give the same grade A or B or C to most of his students and an F to the few others.
a. #Professor Smith gave a B to some of his students.
b. Professor Smith gave a B to most of his students.
c. Every professor gave a B to most of his students.
- (ii) [some/all] Because of an MIT policy against intellectual discrimination, every professor has to give the same grade to all of his students.
a. #Professor Smith gave a B to some of his students.
b. Professor Smith gave a B to all of his students.
c. Every professor gave a B to all of his students.

I am not sure what the judgments are.

The two issues (184) and (185) might admit a common solution. The idea of the common solution is that the strengthened presupposition of the target sentence φ is not that common knowledge entails the negation of the presupposition of the alternative (i.e. $CK(\neg\psi_{prs})$) but rather just that it is not the case that common knowledge entails the presupposition of the alternative (i.e. $\neg CK(\psi_{prs})$). The formalization of this last step is admittedly not straightforward and I defer a more thorough discussion to future work.³⁷

First remark As shown by the examples in (60a)-(64a) and (171a)-(175a), ‘some’ and ‘every’ sound just as odd in the contexts considered. The triplet in (186) summarizes the relevant pattern.

- (186) In a society where children inherit their father’s last name:
- a. #Some children of this couple have a long last name.
 - b. #Every child of this couple has a long last name.
 - c. The children of this couple have a long last name.

I have suggested two slightly different accounts for the oddness of the two cases with ‘some’ and ‘every’: sentence (186a) with ‘some’ is ruled out by the fact that its blind strengthened *meaning* mismatches with common knowledge; sentence (186b) with ‘every’ is ruled out by the fact that its blind strengthened *presupposition* mismatches with common knowledge. But the symmetry between sentences with ‘some’ and sentences with ‘every’ might seem to call for a unified account and thus to cast doubts on my split strategy. As a matter of fact, the following unified account seems readily available, pointed out to me by Danny Fox (p.c) and explored in Spector (2007b). Replace the BH/BH_{prs} with the single hypothesis in (187), which I’ll call the SYMMETRY HYPOTHESIS (henceforth: SH). Furthermore, replace the assumption (176) that ⟨‘the’, ‘every’⟩ is a Horn-scale with the new assumption (188). The pattern in (186) immediately follows: sentences (186a) and (186b) sound odd because they “kill each other” through (187) and (188a); sentence (186c) is fine, because it is spared from (187) by the fact that it has no alternatives according to (188b).

- (187) A sentence sounds odd if it is contextually equivalent to one (all?) of its scalar alternatives.
- (188) a. ⟨‘some’, ‘every’⟩ is a Horn-scale.
 b. ⟨‘the’, ‘every’⟩ and ⟨‘the’, ‘some’⟩ are *not* Horn-scales.

The account based on the SH crucially differs from the account based on the BH/BH_{prs} as follows: the former account is *symmetric*, in the sense that it predicts a sentence to sound odd if it has an alternative contextually equivalent to it; the latter account is *asymmetric*, in the sense that it predicts a sentence to sound odd if it has an alternative logically stronger (but contextually equivalent) to it. I now provide a couple of arguments that the account based on the BH/BH_{prs} is superior to the account based on the SH. My first argument concerns numerals. Example (65), repeated in (189), shows that when two sentences containing numerals are contextually equivalent, the one which is logically stronger sounds fine while the one which is logically weaker sounds odd. The symmetric account based on the SH incorrectly predicts the two sentences in (189) to “kill each other” and thus to both sound odd. The asymmetric account based on the BH makes the right prediction: the blind strengthened meaning of the odd sentence (189a) entails the negation of the corresponding sentence (189b) and thus mismatches with common knowledge; the logically stronger sentence (189b) is fine because its blind strengthened meaning does not mismatch with common knowledge.³⁸ The same point applies to the examples (66) and (67) with disjunction.

- (189) John has an odd number of children. . .
- a. #He has two.
 - b. He has three.

³⁷Sauerland (2003b) advocates a different solution for the problem in (185): he works within the framework of MP (34) and assumes that only alternatives with the same assertion can compete for the sake of MP.

³⁸This argument based on numerals is admittedly confounded by the fact that numerals are known to be rather special scalar items; see Horn (2005) and Geurts (2006) a.o.

Furthermore, consider again the two pairs (99) and (101) studied in the section 2.2. I have shown how the approach based on the BH-MH can account for both these pairs; let me now show that the approach based on the SH cannot. The SH trivially accounts for the oddness of sentence (99a), repeated in (190a): just consider the contextually equivalent alternative (190b), obtained by replacing ‘every day’ with ‘some day’.

- (190) a. #Every day, a fireman won. = (99a)
 b. Some day, a fireman won.

But the SH is not able to account for the fact that sentence (101b), repeated in (191a), sounds fine, given that sentence (191a) is contextually equivalent to the alternative (191b), obtained again by replacing ‘every day’ with ‘some day’.

- (191) a. Every day, for every competition, a fireman won. = (101b)
 b. Some day, for every competition, a fireman won.

Given the crucial role played by the two pairs (99) and (101) for the developments of chapter 3, I stick with the asymmetric account based on the BH/BH_{pr_s} and dismiss the symmetric account based on the SH.³⁹

Second remark Sentences (171a)-(175a) remain odd if the universal quantifier ‘every’ is replaced with the negative quantifier ‘no’, as shown in (192a)-(196a). The extension of the account just suggested for the oddness of sentences (171a)-(175a) to the oddness of sentences (192a)-(196a) would require the assumption that ‘the’ and ‘no’ be Horn-mates, despite the fact that they have different monotonicity.

- (192) In the context described in (171):
 a. φ = #No child(ren) of couple C has a French last name.
 b. ψ = The children of couple C have a French last name.
- (193) Alluding to the Marx Brothers:
 a. #No one’s mother was named Minnie.
 b. Their mother was named Minnie.
- (194) Speaking of the atoms of a molecule which cannot be separated:
 a. #No atom(s) went right.
 b. The atoms went right.
- (195) #No eyes of the victim are blue.
- (196) a. #No Italian comes from a beautiful country.
 b. Italians come from a beautiful country.

³⁹There is one more fact that might speak in favor of the asymmetric account based on the BH/BH_{pr_s} and against the symmetric account based on the SH: despite the fact that both sentences with ‘some’ and ‘every’ in (186) sound odd in contexts which entail homogeneity, they do not sound odd to the same degree. The following example from Chemla (2006) is striking: sentence (ia) with ‘all’ feels way better than sentence (ib) with ‘some’ (italics is not meant to signal focus). Also Singh (2009a) makes the point that oddness in these cases (171a)-(175a) is weaker than in the cases (60a)-(64a).

- (i) In this department, every professor assigns the same grade to his students.
 a. This year, prof. Smith assigned an A to *all* of his students.
 b. #This year, prof. Smith assigned an A to *some* of his students.

This gradience seems hard to account for within the symmetric approach based on the SH: the two sentences (ia) and (ib) should just “kill each other.” Gradience might be easier to account for within the asymmetric approach based on the BH/BH_{pr_s}: sentence (ib) would be ruled out because of a blind mismatching strengthened *meaning* while sentence (ia) would be ruled out because of a blind mismatching strengthened *presupposition*, and the latter mismatch might feel “less harmful” than the former, especially in cases where the relevant common knowledge is only provided by preceding discourse.

A possible way to overcome this difficulty might be the following. Sauerland (2000), building on some existing literature, suggests that the negative quantifier ‘no’ simply does not exist. Rather, it should be analyzed as a morphological realization of negation plus an indefinite. If this hypothesis is on the right track, then I might suggest that scalar alternatives are computed at a level of representation where ‘no’ is still split into negation and an indefinite, and the implicature needed to account for the oddness of sentences (192a)-(196a) is derived by replacing the indefinite with a definite, which have the same monotonicity. I leave the issue open for the time being.

2.5 Miscellaneous issues

2.5.1 An alternative approach based on Manner and its inadequacy

A Reviewer of *Natural Language Semantics* has pointed out to me an alternative account for some of the facts considered in this chapter. The idea of this alternative account has come up also in other presentations of this material. In this subsection, I sketch this alternative account and explain why I consider it inadequate.

Sketch of an approach based on Manner This alternative account shares with the account I have defended in this chapter the rough intuition that a sentence φ sounds odd if there is a sentence ψ such that φ and ψ are contextually equivalent and yet ψ is “better” than φ . The two accounts differ in how they spell out the condition that the alternative ψ be “better” than the target sentence φ . For my account, “better” is inspired to the maxim of Quantity: ψ is better than φ if ψ is logically stronger than φ . For the alternative account, “better” is inspired to the maxim of Manner: ψ is better than φ if ψ is “less costly” than φ , yielding the generalization in (197). The beginning of a definition of the relation “less costly” is provided by the conditions in (198), as suggested by the Reviewer.

- (197) A sentence φ sounds odd if there is a sentence ψ such that φ and ψ are contextually equivalent and furthermore ψ is “less costly” than φ .
- (198) A sentence ψ is less costly than a sentence φ if one of the following conditions holds:
- a. φ contains some more overt words than ψ ;
 - b. φ contains a word of a higher type than the corresponding word in ψ ;
 - c. fronting has been performed in φ but not in ψ .

The alternative account based on (197)-(198) works as follows. Sentence (64a), repeated in (199a), sounds odd because it is contextually equivalent to the alternative (199b) and this alternative is better by (198a), because it does not contain the overt word ‘sometimes’ (rather, it contains a covert generic operator).

- (199) a. #Some Italians come from a beautiful country.
b. Italians come from a beautiful country.

Analogously, sentence (60a), repeated in (200a), sounds odd because it is contextually equivalent to the alternative (200b) and this alternative is better by (198b), because it contains a definite (which has a low type e) rather than the quantifier ‘some’ (which has the higher type $\langle\langle e, t \rangle, t\rangle$).

- (200) In a society where all the children of a given couple have the same last name, namely that of their father:
- a. #Some children of that couple have a Russian last name.
 - b. The children of that couple have a Russian last name.

Finally, sentence (99a), repeated in (201a), sounds odd because it is contextually equivalent to the alternative (201b) and this alternative is better by (198c), because no quantifiers have undergone fronting.

- (201) In the context described in (99):

- a. #On every day, a/some fireman won.
- b. A/Some fireman won on every day.

I'll now argue that this alternative account is not empirically adequate, by challenging one by one the clauses of the definition (198) of "less costly".

Against clause (198a) Clause (198a), together with (197) predicts a sentence φ to sound odd if there is another contextually equivalent sentence ψ which contains less words. I think that this prediction is falsified for instance by (202). Both sentences φ and ψ are perfectly fine in the context considered. Yet, (197)-(198a) predict φ to sound odd, since it is contextually equivalent to ψ (because we know that John's grandmother always bakes three cakes) and contains the extra word 'three'. In other words, assumptions (197)-(198a) do not account for the difference between (202) and, say, (199). Yet, this difference is straightforwardly accounted for by my proposal. Sentence (202a) is of course predicted to be fine, since my proposal compares sentences with respect to logic entailment rather than brevity and sentence (202a) is not contextually equivalent to any logically stronger alternative. Sentence (202b) is contextually equivalent to the logically stronger sentence (202a); yet, numerals and indefinites are plausibly not Horn-mates and thus no mismatching implicature is derived.⁴⁰ The pair in (203) makes the same point.

- (202) John's grandmother loves to bake; whenever she does, she always bakes three identical pies, one for John, one for John's brother and one for John's sister; she likes to try a different recipe every time...
 - a. ... φ = Yesterday, she baked three apple pies.
 - b. ... ψ = Yesterday, she baked apple pies.
- (203) a. Italians are tall.
 - b. Italians are usually tall.

Against clause (198b) Clause (198b) together with (197) predicts a sentence φ to sound odd if there is another contextually equivalent sentence ψ which only differs because it contains a word of a lower type. To test this prediction, consider the following slight variant (204) of my original example (200). In the context considered, where it is known that all children of a given couple inherit the same last name, sentence (204a) sounds to me just as odd as the original sentence (200a). Yet (197)-(198b) predict the variant in (204a) to be fine, contrary to the case of the original sentence (200a); here is why. It is usually assumed that, when a definite is conjoined with another quantifier, its type e is raised to the type $\langle\langle e, t \rangle, t\rangle$, in order to provide conjunction with two constituents whose type "ends in t "; see for instance Partee and Rooth (1983). Of course, clauses (197)-(198b) cannot be checked before type shifting in (204), since the alternative (204b) is uninterpretable before type shifting and thus contextual equivalence between (204a) and (204b) could not be checked. Thus, clauses (197)-(198b) must be checked after type shifting has raised the type of the definite 'the children...' in (204a) from e to $\langle\langle e, t \rangle, t\rangle$. Hence, clause (198b) cannot do any work in the case of (204), since the definite 'the children...' in the alternative (204b) has the same type as the indefinite 'some children...' in the target sentence (204a).

- (204) In a society where all the children of a given couple have the same last name, namely that of their father:
 - a. #A friend of mine and some children of that couple have a Russian last name.
 - b. A friend of mine and the children of that couple have a Russian last name.

⁴⁰ I see two reasons why sentence (202a) does not trigger any offensive implicature, and in particular does not trigger the implicature that the corresponding sentence (202b) is false. One straightforward reason might be that the existential quantifier associated with the object bare plural in (202a) is not a Horn-mate with numerals, hence (202b) is not a scalar alternative of (202a). This assumption seems plausible to me. Alternatively, even if (202b) were indeed a scalar alternative of sentence (202a), it might be the case that (202b) does not end up negated in the strengthened meaning of (202a) because of a "symmetry problem". This second strategy might require proper stipulations on the semantics of plural morphology.

Against clause (198c) Clause (198c) together with (197) predicts a sentence φ to sound odd if it differs from a contextually equivalent sentence ψ only because a certain constituent has undergone fronting in φ . To test this prediction, consider again the example (101), repeated in (205). In the context considered, the two sentences (206a) and (206b) are contextually equivalent to the two sentences (205a) and (205b), respectively. Note that both sentences (205a) and (205b) differ from the corresponding variants in (206a) and (206b) because of fronting of the phrase ‘every day’. Thus, (197)-(198c) predict both sentences (205) to sound odd, since both admit a contextually equivalent sentence with no fronting. This prediction is wrong, since sentence *refex: viiib*) sounds fine in the context considered.

- (205) In the context described in (101):
- a. #Every day, a/some fireman won the running competition.
 - b. Every day, for every competition, a/some fireman won.
- (206) a. Some fireman won the running competition on every day.
 b. For every competition, some fireman won it on every day.

I conclude that the current definition (198) of the relation “less costly” is not consistent with the data. I do not see any way to improve on the current definition (198). I therefore dismiss this alternative account inspired to the maxim of Manner.

2.5.2 Predecessors

Some of the ideas presented in this chapter have various predecessors in the recent literature, besides Hawkins’ work mentioned in chapter 1. In this subsection, I mention two such predecessors.

Presupposition accomodation The mandatoriness problem discussed in subsection 2.1.3 has an important precedent in the literature on presupposition accomodation; see for instance Gazdar (1979). Let me illustrate this important precedent with an example. A sentence such as (207) can be uttered felicitously out of the blue. This means that the presupposition that John used to drink beer on Mondays and Tuesdays can be accommodated.

- (207) John has stopped drinking beer every Monday and Tuesday night.

Let’s now embed this sentence (207) in a piece of text such as (208). This text sounds infelicitous: accommodation of the presupposition of sentence (208b) is somehow blocked. This is surprising: if in the case of (207) we are able to accommodate both the fact that John drunk beer on Mondays and the fact that he drunk beer on Tuesdays, then we should *a fortiori* be able to accommodate the single fact that John drunk beer on Tuesdays in the case of (208b). What blocks accommodation in the case of (208)? Various authors have suggested the following answer: sentence (208a) triggers the implicature that the speaker doesn’t know whether John used to drink beer on days of the week that are not Mondays; but sentence (208b) presupposes that John used to drink beer on both Mondays and Tuesdays; thus, the oddness of the text (208) follows from the mismatch between the implicature of (208a) and the requirement on common knowledge imposed by the presupposition of (208b). Note that this is exactly Hawkins’ line of reasoning, as exemplified in (22), (45) and (55). Thus, this line of account for the oddness of (208) runs right away into the Mandatoriness issue (25)/(27): why is it that the implicature is kept in place and wins over the presupposition?

- (208) a. John used to drink beer every Monday night. . .
 b. . . ?but he has stopped drinking beer every Monday and Tuesday night.

The issue can be sharpened further by embedding the text (208) into the dialogue (209), which mimics the one in (84). Again, I think that accommodation of the presupposition of (209b) is not straightforward in this case. Again, we might want to suggest that accomodation of the presupposition of (209b) is challenged by the implicature triggered by (209a). But in the case of (209), B is answering a question concerning who had the habit of drinking beer on Mondays. As noted above for the case in (84), such a question provides a context in which the implicature of (209a) that

drinking only happened on Mondays should not be relevant and should thus be missing. But if it is not relevant and thus missing, then this implicature cannot mismatch with the presupposition of (209b). In conclusion, we have to assume that, although the implicature triggered by (209a) is not in general relevant when uttered as an answer to A's question, nonetheless it is relevant when it triggers a mismatch.

- (209) A: Which one of your friends used to drink beer every monday night?
 B: a. John used to drink beer every monday night...
 b. ...?but he has stopped drinking beer every monday and tuesday night.

The proposal made in section 2.1.3 *might* extend to the cases (208) and (209). Let me quickly discuss this issue. Let φ be a shorthand for the sentence in (208a) or (209a), as stated in (210a). Assume that ψ in (210b) is a scalar alternative of φ in (210a). Thus, the strengthened meaning $\text{EXH}(\varphi)$ of φ is the one in (210), that negates that ψ is both true and relevant. In order to accommodate the presupposition of (208b) or (209b), we need to revise common knowledge \mathcal{W}_{ck} in such a way that it entails that John drank beer on Tuesdays. Given such a common knowledge, the two alternatives φ and ψ in (210) are contextually equivalent. Since φ is relevant by (89a) and since relevance is closed w.r.t. contextual equivalence by (89b), then ψ is relevant too. Thus, the strengthened meaning in (210) entails the negation of ψ and thus mismatches with the common knowledge needed in order to accommodate the presupposition.⁴¹

- (210) $\text{EXH}(\varphi) = \varphi \wedge (\neg\psi \vee \neg\mathcal{R}(\psi))$
 a. $\varphi =$ John drinks on Mondays.
 b. $\psi =$ John drinks on Mondays and Tuesdays.

Fox and Hackl (2006) They offer an argument in favor of the BH (82), which I review quickly here.⁴² As a starting point, consider the facts in (211): the strengthened meaning of sentence (211a) is intuitively equivalent to the plain meaning of sentence (211b), namely to the proposition that John weighs *exactly* 120 pounds. These facts are very easy to account for, as in (212). Let me denote by φ the sentence (211a) and the prejacent of 'only' in sentence (211b), as in (212a). Let me make the assumption that '120' is a Horn-mate of the degree 'd', for any real number $d \in \mathbb{R}$. As stated in (212b), the set of excludable alternatives then consists of all the alternatives ψ_d of the form 'John weighs d pounds' where d is a degree strictly larger than 120. This characterization of the set of excludable alternatives holds no matter whether it is computed according to logic entailment (76a) or according to entailment relative to common knowledge (76b), since in this case there is no relevant piece of common knowledge to take into account. The intuitively correct result in (212c) is thus straightforwardly derived.

⁴¹I see two potential problems with this line of reasoning. The *first* problem concerns timing of accommodation. In order for the account just sketched to work, it is crucial that the presupposition of (208b) or (209b) must be accommodated before the strengthened meaning (210) of (208a) and (209a) is computed, namely that the two sentences in the text must be treated as piece and parcel. In fact, if instead the strengthened meaning (210) were computed before the presupposition were accommodated, then the alternatives φ and ψ would not count as contextually equivalent (since it is the accommodated presupposition that makes them contextually equivalent) and the implicature $\neg\psi$ would thus fail to be mandatory. The *second* problem concerns whether ψ in (210b) is the proper alternative or whether it should be replaced by ψ' in (1).

(1) $\psi' =$ John drinks on Tuesdays.

Let me make the issue explicit. On the one hand, it is crucial to use ψ instead of ψ' in the preceding account. In fact, the piece of common knowledge that John drank on Tuesdays renders φ and ψ contextually equivalent but it does not render φ and ψ' contextually equivalent. On the other hand, it is not obvious that ψ is a proper alternative of φ . Alternatives such as ψ are common in the literature on 'only', to get right the semantics of sentences such as 'Only John came' on the background of the standard definition (75) of the set of excludable alternatives in terms of asymmetric entailment. But in section 2.2, I followed Fox (2007a) and replaced the standard definition (75) with the variant (106) that drops the requirement of asymmetric entailment. From the perspective of this latter definition, I think the alternative ψ' makes more sense than the alternative ψ . Furthermore, the alternative ψ' but not the alternative ψ is compatible with the intuition that alternatives cannot be syntactically more complex than the target sentence, as spelled out for example in Katzir (2008). This second problem is analogous to the one discussed in the Second Remark in section 2.2.

⁴²Fox and Hackl extend the ideas summarized here to the case of negative islands. See also Fox (2007b), Spector (2006), Abrusan (2009) and Abrusan and Spector (2009) for discussion of this extension.

- (211) a. John weighs 120 pounds.
 b. John weighs only 120_F pounds.
- (212) a. $\varphi =$ John weighs 120 pounds.
 b. $\mathcal{E}xcl(\varphi) = \{\psi_d \mid d > 120\}$, where $\psi_n =$ John weighs (at least) n pounds.
 c. $\text{EXH}(\varphi) = \llbracket \text{only} \rrbracket(\varphi) =$ John weighs exactly 120 pounds.

Consider next the facts in (213), obtained from (211) by adding ‘more’. Sentence (213a) does not trigger any implicature, namely does not mean that John weighs exactly 121 pounds. Furthermore, sentence (213b) with overt ‘only’ sounds deviant. Here is a way to account for these facts (213). Let me denote by φ the sentence (213a) and the prejacent of ‘only’ in sentence (213b), as in (214a). As stated in (214b), in this case the set of excludable alternatives is empty. In fact, any alternative ψ_d with $d \leq 120$ is not innocently excludable given φ , according to Fox’s definition (106). Furthermore, the set $\{\psi_d \mid d > 120\}$ is not innocently excludable either, since $\varphi \wedge \bigwedge \{\neg\psi_d \mid d > 120\}$ is a logical contradiction, and thus a fortiori a contradiction given common knowledge. Finally, all proper subsets of the set $\{\psi_d \mid d > 120\}$ are innocently excludable, but their intersection is empty. The characterization (214b) of the set of excludable alternatives holds no matter whether it is computed according to logic entailment (76a) or according to entailment relative to common knowledge (76b), since in this case there is no relevant piece of common knowledge to take into account. As stated in (214c), strengthening is correctly predicted to be vacuous. The deviance of sentence (213b) with overt ‘only’ is expected if there is indeed some constraint that rules out vacuous instances of overt ‘only’, as advocated above for the case of sentence (80).⁴³

- (213) a. John weighs more than 120 pounds.
 b. ?John weighs only more than 120_F pounds.
- (214) a. $\varphi =$ John weighs more than 120 pounds.
 b. $\mathcal{E}xcl(\varphi) = \emptyset$
 c. $\text{EXH}(\varphi) = \llbracket \text{only} \rrbracket(\varphi) = \varphi$.

The crucial observation by Fox and Hackl is that (215) is completely analogous to (213): sentence (215a) does not trigger any implicature; and sentence (215b) with overt ‘only’ sounds deviant. Fox and Hackl note that, in order to maintain the parallelism between (215) and (213), it is crucial to assume that the computation of the set of excludable alternatives in the case of (215) is blind to the piece of common knowledge that children come in wholes, contrary to pounds.

- (215) a. John has more than 3 children.
 b. ?John only has more than 3_F children.

Here are the details. If we compute the the set of excludable alternatives using logic entailment (76a), then there is no difference at all between children and pounds. Thus, we correctly predict there to be no difference between the two cases (215) and (213), as noted in (216).

- (216) Using logic entailment (76a):
 a. $\varphi =$ John has more than 3 children.
 b. $\mathcal{E}xcl(\varphi) = \emptyset$
 c. $\text{EXH}(\varphi) = \llbracket \text{only} \rrbracket(\varphi) = \varphi$.

As noted in (217), things are very different if we use entailment relative to common knowledge (76b). In this case, φ in (217a) and the alternatives ψ_d ’s obtained from φ by replacing ‘3’ by ‘ d ’, are contextually equivalent if $3 \geq d > 4$, because to have at least 3.5 children is equivalent to have at least 4 children, since children come in wholes. Thus, the alternatives ψ_d ’s with $3 \geq d > 4$ are not excludable w.r.t. φ with respect to entailment relative to common knowledge and the maximal set

⁴³This line of reasoning would predict sentence (213b) to improve if ‘only’ were to associate with something else, besides the numeral.

of excludable alternatives contains all the alternatives ψ_d with $d > 4$. As noted in (217c), we thus incorrectly predict the strengthened meaning of sentence (215a) to say that John has exactly four kids; furthermore, we predict the overt occurrence of ‘only’ in sentence (215b) not to be vacuous, leaving the deviance of sentence (215b) unaccounted for.

- (217) Using contextual entailment (76b):
- a. $\varphi =$ John has more than 3 children.
 - b. $\mathcal{E}xcl(\varphi) = \{\psi_d \mid d > 4\}$
 - c. $\mathcal{E}XH(\varphi) = [\text{only}](\varphi) =$ John has exactly 4 children.

2.5.3 Miscellaneous problematic cases

This final subsection collects a few miscellaneous problematic cases. They do not seem to have anything in common, apart from the fact that I do not quite know what to say about them.

First case Singh (2009b) notes the oddness of sentence (218b). This case is not accounted for by the proposal developed in this chapter. In fact, the constituent ‘ n or more’ does not trigger any scalar implicature, as shown by the case of sentence (218a); see for instance Chierchia et al. (to appear) for a discussion of these type of sentences. Since the strengthened meaning of sentence (218b) is thus identical to its plain meaning, the oddness of sentence (218b) cannot be due to any mismatching implicature.

- (218) a. John has two or more sons.
b. #I have two or more sons.

Given the proposal developed in this section, it might be somewhat natural to suggest the following line of account. Sentence (218a) does trigger the inference that the speaker does not know whether John has exactly two sons or whether he has more than that. After Sauerland (2004b), this type of ignorance inferences are called PRIMARY IMPLICATURES. If sentence (218b) triggers the same primary implicature, then its oddness might be due to the fact that this primary implicature mismatches with the piece of common knowledge that people know how many children they have. Of course, this line of reasoning would require primary implicatures to be computed blind to common knowledge and to be mandatory, in order to overcome the two issues (25)/(27) and (26)/(28).

Second case Spector (2007b) notes the oddness of sentence (219b): somehow, the sentence mismatches with the piece of common knowledge that people can marry only one person (at the time). This case is not accounted for by the proposal developed in this section. In fact, even if sentence (219b) were to trigger the implicature that John did not marry *two* Italian women, this implicature would not trigger any mismatch with common knowledge. The variant in (219c) shows that the problem extends to the case with overt ‘only’.

- (219) a. Last summer, John married an Italian woman.
b. #Last summer, John married one Italian woman.
c. #Last summer, John only married one_F Italian woman.

Third case In section 2.1, I have considered sentences such as (64a), repeated below in (220a), and I have accounted for its oddness as follows: its blind strengthened meaning says that “some but not all Italians come from a beautiful country”, which of course mismatches with the piece of common knowledge that all Italians come from the same country. Unfortunately, this account does *not* extend to the variant in (220b), which feels just as odd: the blind strengthened meaning of (220b) would be “some but not all Italians come from a beautiful country and are blond”, which does not contradict common knowledge, since it would be true in worlds where not all Italians are blond. The pair in (221) makes the same point.

- (220) a. #Some Italians come from a beautiful country.
 b. #Some Italians come from a beautiful country and are blond.
- (221) In the context described in (171):
 a. #Some children of couple C have a French last name.
 b. #Some children of couple C have a French last name and are blond.

To fully appreciate the problem, consider (222), which is obtained from (220b) by replacing 'some' by 'all' and furthermore 'and' by the operator LEFT of Sauerland (2004c). If (222) were a scalar alternative of (220b), then the problem would be solved, since the blind strengthened meaning of (220b) would then entail the negation of (222) and thus mismatch with common knowledge. But (222) is not a scalar alternative of (220b) according to Fox's new definition (124): the replacement of 'and' by LEFT leads to a weaker formula, no matter whether it is performed before or after the replacement of 'some' by 'all'.⁴⁴

(222) All Italians come from a beautiful country.

Let me also note that the facts are not as sharp as one might expect. The following cases are formally parallel to those in (220)-(221), and yet the (b) sentences with conjunction sound quite fine to me.

- (223) John gave the same grade to all of his students. . .
 a. #He gave an A to some of them.
 #To some of them, he gave an A.
 b. He gave an A and a prize to some of them.
 To some of them, he gave an A and a prize.
- (224) John always reads the same newspaper for a month and then switches to another newspaper the next month. During this month. . .
 a. #John sometimes read the New York Times.
 b. John sometimes read the New York Times and drank a cup of coffee.

I leave the issue open for the time being.

⁴⁴ Note that the problem could be overcome by repeating the same account at the level of presupposition; let me illustrate. Let me use φ as a shorthand for sentence (220b), as stated in (ia); let ψ be sentence in (ib), obtained from φ by replacing 'some' by the generic operator GEN and furthermore 'and' by LEFT. As stated in (iia), sentence φ in (ia) has vacuous presuppositions. Following von Stechow (1997), assume that GEN has a homogeneity presupposition (see (435) below for an explicit definition); thus, the alternative ψ in (ib) has the presupposition in (iib).

- (i) a. $\varphi = ?$ Some Italians come from a beautiful country and are blond.
 b. $\psi =$ GEN Italians come from a beautiful country.
- (ii) a. $\varphi_{\text{prs}} = \mathcal{V}$.
 b. $\psi_{\text{prs}} =$ Either all or no Italians come from a beautiful country.

If the replacement of 'and' by LEFT is performed before the replacement of 'some' by GEN, then it leads to an alternative whose presupposition is not weaker than the presupposition (iia) of φ (namely, it is just as vacuous as that of φ) and the replacement is thus licit according to Fox's new definition (124). The strengthened presupposition of φ would thus entail $\neg\psi_{\text{prs}}$ and therefore mismatch with the common knowledge that all Italians come from the same country.

Chapter 3

Application to individual level predicates

As reviewed in section 1.1, the s-/i-predicates distinction has many grammatical reflexes. Thus, it looks like the distinction should be grammatically encoded. Various ways to encode this distinction into grammar have been suggested in the literature. I have provided a quick sample in (41), repeated below in (225). In this chapter, I will discuss in some detail these assumptions (225) and various other similar assumptions.

- (225)
- Kratzer (1995)*: an i-predicate lacks a Davidsonian argument, contrary to s-predicates.
 - Chierchia (1995)*: an i-predicate bears a feature that forces local agreement with a covert generic operator that mandatorily binds the Davidsonian argument of the i-predicate.
 - Diesing (1992)*: an i-predicate is selected by a special inflectional head that requires its subject to be base generated in [Spec, IP], rather than in [Spec, VP] as it is the case for s-predicates.

Departing from these and other similar grammatical characterizations of i-predicates, in this chapter I will assume that there is no relevant grammatical difference between s- and i-predicates: there is no difference with respect to the position where their subjects are base-generated; there is no difference with respect to their syntactic features; and there is no difference with respect to their argumental structure, as stated in (226).

- (226) Both the i-predicate ‘tall’ and the s-predicate ‘available’ have a Davidsonian argument. For simplicity, I will naïvely take that argument to range over times $t \in \mathcal{T}$.

What is the relevant difference between s- and i-predicates, then? I assume that all that’s special about an i-predicate such as ‘tall’ is that it follows from common knowledge \mathcal{W}_{ck} that, if an individual is tall at a given time, then he is tall throughout his entire life span. Let me state this assumption as in (227), where $\lambda t . \text{in}^w(d, t)$ stands for the life span of an individual d at a world w .

- (227) For every individual $d \in \mathcal{D}_e$ and for every world $w \in \mathcal{W}_{ck}$ compatible with common knowledge: if there exists a time $t' \in \mathcal{T}$ such that $[[\text{tall}]]^w(d, t')$, then $[[\text{tall}]]^w(d, t)$ for every time t such that $\text{in}^w(d, t)$.

As noted in section 1.1, s-predicates can do many more things than i-predicates can do: there are configurations where s-predicates are fine while i-predicates are not; and there are readings that are available with s-predicates but not with i-predicates. A theory of i-predicates should thus account for why i-predicates cannot do the many things that s-predicates can do. In this chapter, I will argue that, under assumptions (226)-(227), such a theory is a theorem of the more general theory of oddness sketched in chapter 2, as roughly stated in (228). As a matter of fact, I will argue that, under assumptions (226)-(227), various properties of i-predicates are formally analogous to various cases considered in chapter 2, that had nothing to do with i-predicates.

- (228) Those many things that i-predicates cannot do would correspond to a blind strengthened meaning (or a blind strengthened presupposition) that contradicts the piece of common knowledge in (227).

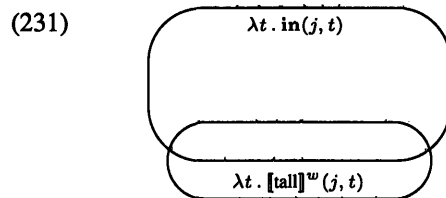
Before I move on to argue in favor of (228), let me comment on the core assumptions (226) and (227) with a few remarks. Assumption (226) is of course an oversimplification, since it construes the Davidsonian argument in too simple a way, namely just as a time, rather than as a full event or state or situation. I stick to this simplification only to keep my ontological assumptions on davidsonian arguments at a minimum. Of course, once assumption (226) is restated using, say, events or states, then it can be tested using the standard battery of arguments for davidsonian arguments. For instance, Landman (2000, Ch. 1) notes that sentence (229a) entails the other sentences in (229) and thus concludes that “in as much as the modifier argument is evidence for an event variable for s-predicates, it is evidence for a state variable for i-predicates”. I will briefly come back to this issue in section 3.1, where I’ll discuss Kratzer’s (1995) proposal that i-predicates lack the Davidsonian argument.

- (229) a. I know John well by face from TV.
 b. I know John by face from TV.
 c. I know John well by face.
 d. I know John by face.

Assumption (227) captures the intuition that i-predicates denote permanent properties quite crudely. Of course, I don’t really want to say that, if John is tall at a given time, then he has got to be tall at literally *every* time t throughout his life span: tall men might have been short kids. A more careful restatement of assumption (227) should thus replace John’s whole life span (230a) with some proper subset as in (230b), which might depend on the specific i-predicate considered, be vaguely defined and context dependent. In what follows, I will ignore this kind of complications and stick to the simple formulation (227) in terms of (230a). As far as I can see, nothing in my proposal hinges on this simplification. I will briefly come back to this issue in section 3.2.

- (230) a. $\lambda t . \text{in}^w(j, t)$
 b. $\lambda t . \text{C}_{\text{tall}}^w(j, t)$

An equivalent way of stating assumption (227) is as follows: in every world compatible with the piece of common knowledge (227), the intersection between the lifespan $\lambda t . \text{in}^w(j, t)$ of John and the set of times $\lambda t . \llbracket \text{tall} \rrbracket^w(j, t)$ at which he happens to be tall is either empty or else coincides with the entire life span $\lambda t . \text{in}(j, t)$. In other words, the predicate $\lambda t . \llbracket \text{tall} \rrbracket(j, t)$ is HOMOGENEOUS w.r.t. the restrictor $\lambda t . \text{in}(j, t)$, in the sense of section 1.3: the situation depicted in (231)¹, where John is tall at some times in his life span but not at others, can never arise in worlds compatible with common knowledge.



Another equivalent way of stating assumption (227) is as follows: the two propositions in (232a) are equivalent given the piece of common knowledge (227), even though ψ *asymmetrically* logically entails φ . Assumption (227) is stated for the case of the i-predicate ‘tall’ but it trivially extends to all i-predicates with a single individual argument. Its extension to the case of transitive i-predicates, such as ‘know’, ‘love’ or ‘hate’, is more delicate. Again, it should be stated in such a way that the two propositions in (232b) are contextually equivalent. The issue in this case is how to define the restrictor C : it seems plausible that it should somehow depend on both the life spans of the subject

¹With the caveat of footnote 9.

John and the direct object Mary; but it is not obvious whether they should play a symmetrical role or not.

- (232) a. $\varphi = \lambda w . \exists_t [\text{in}^w(\text{john}, t)] [[\text{tall}]^w(\text{john}, t)]$.
 $\psi = \lambda w . \forall_t [\text{in}^w(\text{john}, t)] [[\text{tall}]^w(\text{john}, t)]$.
 b. $\varphi = \lambda w . \exists_t [\text{C}^w(\text{john}, \text{mary}, t)] [[\text{love}]^w(\text{john}, \text{mary}, t)]$.
 $\psi = \lambda w . \forall_t [\text{C}^w(\text{john}, \text{mary}, t)] [[\text{love}]^w(\text{john}, \text{mary}, t)]$.

Finally, note the crucial difference between the characterization of i-predicates in (227) and those quoted in (225) from the recent literature. The latter characterizations are grammatical in nature. My characterization (227) has nothing to do with grammar; rather, it is an assumption on common knowledge \mathcal{W}_{ck} . As far as grammar is concerned, there is no difference between s- and i-predicates: there are possible worlds where the i-predicate ‘tall’ denotes a permanent property, in the sense of (227); and there are other possible worlds where it denotes a non-permanent property, just as is the case for the s-predicate ‘available’. The only difference between the two predicates is that \mathcal{W}_{ck} contains no worlds where the extension of ‘tall’ is non-permanent but does contain worlds where the extension of ‘available’ is non-permanent.

3.1 Existential Q-adverbs

As shown in (233), i-predicates do not tolerate existential Q-adverbs, contrary to s-predicates. As shown in (234), the contrast disappears once the definite subject ‘John’ is replaced by an indefinite or a bare plural. These same facts hold with Q-adverbs of arbitrary quantificational force, such as ‘always’, ‘often’ or ‘never’. In this section, I concentrate on the case of adverbs with existential quantificational force, such as ‘sometimes’; in section 3.7, I will come back to the case of Q-adverbs with universal or generic quantificational force.

- (233) a. Sometimes, John is available.
 b. #Sometimes, John is tall.
- (234) a. Sometimes, firemen are available.
 b. Sometimes, firemen are tall.

In this section, I argue that the oddness of sentence (233b) can be accounted for in exactly the same way I have accounted in section 2.1 for the oddness of sentences (60a)-(64a), one of which is repeated in (235).

- (235) #Some Italians come from a beautiful country.

My argument exploits the following parallelism: in the case of sentence (233b), assumption (227) ensures that it follows from common knowledge \mathcal{W}_{ck} that, if John is sometimes tall, then he always is; analogously, in the case of sentence (235), it follows from the common knowledge that all Italians come from the same country, that if some Italians come from a beautiful country, then they all do. In order to stress the analogy between the account for the oddness of sentence (235) and that for the oddness of sentence (233b), I summarize the main steps of the two accounts one next to the other in Table 3.2.

3.1.1 Existing accounts

Two main lines of analysis have been suggested in the literature for the facts (233) and (234) — as well as for the analogous facts with other Q-adverbs. I discuss them in this subsection, trying to argue that they are not fully satisfactory.

#Some Italians come from a beautiful country.	#John is sometimes tall.																
(1) LOGICAL FORM																	
By assumption (86), the LF contains a mandatory exhaustivity operator																	
(2) TRUTH CONDITIONS OF THE PREJACENT																	
$\varphi = \exists_x [[\text{Italians}](x)] [[\text{VP}](x)]$	$\varphi = \exists_t [C^w(t) \wedge \text{in}(j, t)] [[\text{tall}](j, t)]$																
(3) ASSUMPTION ON HORN SCALES																	
⟨‘some Italians’, ‘all Italians’⟩	⟨‘sometimes’, ‘always’⟩																
(4) SET OF SCALAR ALTERNATIVES																	
$\text{Alt}(\varphi) = \{\psi\}$, where ψ is the LF with the following shape and truth conditions																	
$\psi = \forall_x [[\text{Italians}](x)] [[\text{VP}](x)]$	$\psi = \forall_t [C(t) \wedge \text{in}^w(j, t)] [[\text{tall}](j, t)]$																
(5) SET OF EXCLUDABLE ALTERNATIVES																	
$\text{Excl}(\varphi) = \{\psi\}$, since $\varphi \wedge \neg\psi$ is true e.g. in the following world:																	
<table border="1" style="border-collapse: collapse; margin: auto;"> <tr> <td></td> <td style="text-align: center;">Aldo</td> <td style="text-align: center;">Maria</td> <td style="text-align: center;">Anna</td> </tr> <tr> <td style="text-align: center;">beautiful:</td> <td style="text-align: center;">✓</td> <td></td> <td style="text-align: center;">✓</td> </tr> <tr> <td style="text-align: center;">not-beautiful:</td> <td></td> <td style="text-align: center;">✓</td> <td></td> </tr> </table>		Aldo	Maria	Anna	beautiful:	✓		✓	not-beautiful:		✓		<table style="margin: auto;"> <tr> <td style="text-align: center;">$C^w(\cdot) \wedge \text{in}^w(j, \cdot)$</td> </tr> <tr> <td style="text-align: center;">—————</td> </tr> <tr> <td style="text-align: center;">$[[\text{tall}]]^w(j, \cdot)$</td> </tr> <tr> <td style="text-align: center;">—————</td> </tr> </table>	$C^w(\cdot) \wedge \text{in}^w(j, \cdot)$	—————	$[[\text{tall}]]^w(j, \cdot)$	—————
	Aldo	Maria	Anna														
beautiful:	✓		✓														
not-beautiful:		✓															
$C^w(\cdot) \wedge \text{in}^w(j, \cdot)$																	
—————																	
$[[\text{tall}]]^w(j, \cdot)$																	
—————																	
(6) DOMAIN Ω OF THE EXHAUSTIVITY OPERATOR																	
$\Omega = \{\psi\}$, by assumptions (89)																	
together with the fact that φ and ψ are contextually equivalent																	
(7) CONCLUSION																	
the sentences denote the proposition $\text{EXH}_\Omega(\varphi) = \varphi \wedge \neg\psi$ that contradicts common knowledge; oddness thus follows from the MH (97)																	

Table 3.2: Parallelism between the account for the oddness of sentence (64a) presented in section 2.1 and the account for the oddness for sentence (233b) presented in this section.

Kratzer (1995) and variants thereof Kratzer (1995) doesn't explicitly discuss the case of sentences containing 'sometimes' and an i-predicate; she only discusses the case of 'when'-clauses. But her account for 'when'-clauses extends straightforwardly to the case of 'sometimes', as follows. Kratzer's core assumption is (236).

(236) I-predicates lack a Davidsonian argument, contrary to s-predicates.

The deviance of sentence (233b) can then be straightforwardly accounted as follows: by assumption (236), sentence (233b) ends up with the truth conditions (237a); the adverb 'sometimes' has nothing to quantify over in the case of these truth conditions; this case is thus ruled out by a constraint against vacuous quantification; see Potts (2002) for a discussion of this constraint. The case of (234b) is very different: following a large body of literature, Kratzer assumes that the corresponding truth conditions are (237b), where the bare plural 'firemen' introduces a variable x that gets quantified over by the adverb 'sometimes' (see section 3.3 for more details); no vacuous quantification arises in this case.

- (237) a. $\text{SOMETIMES}_t[[\text{tall}]](j)$
 b. $\text{SOMETIMES}_x[[\text{fireman}]](x) \wedge [\text{tall}]](x)$

Kratzer's assumption (236) has been criticized by a number of authors, based on the observation that non-temporal adverbial modification is indeed possible with i-predicates, thus requiring an event argument also in the case of i-predicates. I repeat in (238) the examples in (229), from Landman (2000, pp. 15-17).²

- (238) a. I know John well by face from TV.
 b. I know John by face from TV.
 c. I know John well by face.
 d. I know John by face.

In the face of the data in (238), Kratzer's initial assumption (236) can be rather trivially modified: instead of assuming that i-predicates lack a Davidsonian argument altogether, we can assume that they do have a Davidsonian argument too, but that their Davidsonian argument is special in not being "spatio-temporally located", as suggested for instance in Jäger (2001).³ For concreteness, here is a possible implementation of this idea. Assume that all predicates, both s- and i-predicates, have an event argument. Assume furthermore that the temporal dimension of an event is recovered through a function τ that maps an event e to its corresponding time interval $\tau(e)$. We could then replace Kratzer's strong assumption (236) with something like the weaker assumption (239). Since i-predicates do have Davidsonian arguments, the challenge raised by the instances of adverbial modification in (238) is immediately met, since these adverbs do not modify the time dimension of events. But since 'sometimes' wants to quantify over the time dimension of events, then Kratzer's account for the pattern in (233) and (234) remains essentially unaltered.⁴

²All i-predicates appear to be statives. Kratzer's assumption (236) has later been extended by some authors from i-predicates to statives in general; see for instance Katz (2003); I will come back to the relationship between i-predicates and statives in subsection 3.6.4. It seems to me that authors that have more recently endorsed some version of Kratzer's assumption (236) have failed to cope with the challenge raised by instances of adverbial modifications such as those in (238). For instance, Katz (2003) seems to suggest that all adverbial modifiers with statives should be interpreted as "predicate modifiers and not as event predicates." But the distinction between "predicate modifiers" and "event predicates" seems suspicious to me. No tests are provided in order to draw the line. Furthermore, once we allow for "predicate modifiers", then the very crucial argument for events based on adverbial modification evaporates.

³See footnote 26 below for a more detail discussion of Jäger's (2001) position on the issue.

⁴Here is a way of spelling out the details. Consider first the case where 'sometimes' in sentence (233b) is quantifying over events, as in the truth conditions (ia). In this case, there is no vacuous quantification, and thus something else is needed in order to rule out these truth conditions. For instance, assume that 'sometimes' carries a presupposition against events that are not temporally located, as in (iii). Truth conditions (ia) are thus immediately ruled out by (iii) and (239), since they suffer from presupposition failure.

- (i) a. $\text{SOMETIMES}_e[[\text{tall}]](j, e)$
 b. $\exists_e \text{SOMETIMES}_\tau[[\text{tall}]](j, e)$

- (239) The function τ is partial and its domain $dom(\tau)$ does not contain any event in the extension of an i-predicate.

Yet, I think that this weaker assumption (239) can be questioned too, for instance by means of the following argument.⁵ It is well known that i-predicates can be coerced into s-predicates. According to Kratzer's original assumption (236), this means that the English lexical entry 'tall' is ambiguous between an i-predicate exponent without the Davidsonian argument and an s-predicate exponent with the Davidsonian argument. According to the weaker assumption (239), this means that the English lexical entry 'tall' is ambiguous between an i-predicate exponent with a non-temporally-located Davidsonian argument and an s-predicate exponent with a temporally located Davidsonian argument. The crucial point here is that, once we grammaticalize the s- vs. i-predicates distinction by means of assumptions such as (236) or (239), coercion requires lexical ambiguity. But then, the contrast in (240) is left unexplained: coercion of an i-predicate into an s-predicate would correspond to a special kind of lexical ambiguity, that patterns differently from plain cases of lexical ambiguity.⁶

- (240) a. #John went to the bank to get money, and Bill did too to get some water.
b. John is intelligent. And Mary is too, when she wants to be.

De Swart (1991) Departing from this line of accounts based on (236) or (239), de Swart (1991, pp. 62-67) makes a rather different proposal, that is very close in spirit to the one I am defending in this work. She notes that predicates such as 'to die', 'to kill' or 'to build John's house' behave very similarly to i-predicates with respect to adverbial modification, as illustrated by the pairs in (241)-(242). What is special about these predicates is that they are not iterable, namely they can hold only once, so to speak. Thus, de Swart calls them ONCE-ONLY PREDICATES. Because of the parallelism in (241)-(242), de Swart suggests that all that's special about i-predicates is that they are once-only s-predicates: "What i-predicates and once-only predicates have in common is that their application to a particular individual is felicitous only once. [...] The set of spatio-temporal locations that is associated with an i-predicate or a once-only predicate is a singleton set for all models and each assignment of individuals to the arguments of the predicate" (p. 65). Chierchia (1995) adopts this proposal too.⁷

- (ii) a. $SOMETIMES_{e,x}[[\text{fireman}]](x) \wedge [\text{tall}](x, e)$
b. $\exists_e SOMETIMES_x[[\text{fireman}]](x) \wedge [\text{tall}](x, e)$
(iii) $SOMETIMES_e = \lambda X_{\langle s,t \rangle} : X \subset dom(\tau) . X \neq \emptyset$

Some care is needed in order to make this strategy compatible with the felicity of sentence (234b): of course, we cannot posit truth conditions (iia) for this sentence, since these truth conditions suffer from presupposition failure just as those in (ia). Thus, in this case we would need to assume that 'sometimes' is allowed to quantify only over firemen, while a covert operator \exists_e is taking care of the event argument, as in truth conditions (iib). Positing this covert operator over eventualities of course does not threaten the initial account for the deviance of sentence (233b): positing the null operator for the case of this sentence leads to truth conditions (ib), where 'sometimes' suffers from vacuous quantification. Note finally that, even if Kratzer's original account for the pattern in (233) and (234) is not substantially affected by the replacement of her strong assumption (236) with the weak variant (239), other consequences that she derives from her strong assumption (236) do not seem to me to follow anymore from the weaker variant (239). This seems to me to be the case in particular for her derivation of Diesing's generalization; see footnote 22.

⁵The argument seems rather straightforward to me, but I have not seen it in the literature yet.

⁶I have to concede that sentence (240b) might be problematic for my account too. To appreciate the problem, consider first a plain case of coercion such as (i). According to my proposal, coercion in (i) can be treated without lexical ambiguity. Rather, coercion in (i) might be construed as a process of revision of common knowledge (227), in such a way that it does not follow anymore from common knowledge that if Mary is intelligent, then she always is.

(i) Mary is intelligent when she wants to be.

Yet, two problems immediately arise, that I will leave open for the time being. One problem is that a proper formalization of this intuition would require a proper, fine-grained restatement of the MH (97), that is able to discriminate cases where contextual mismatch leads to coercion from cases where it leads to oddness. Another problem is that this revision of common knowledge (227) must be construed carefully, so as to only apply to Mary but not to John in the case of sentence (240b). Perhaps, some thing like the idea that coercion consists of a *minimal* modification of common knowledge (227) might be made to work here.

⁷Chierchia does not spell out in detail how de Swart's proposal can be casted within his framework. He just writes: "[...] It seems plausible to maintain that [...] variables [...] must in principle be satisfiable by more than one entity. Let's call

- (241) a. #John is always / sometimes / often tall.
 b. #John always / sometimes / often kills Mary.
- (242) a. A fireman is always / sometimes / often tall.
 b. John always / sometimes / often kills a rabbit.

Yet, note that once-only predicates seem to be all either accomplishments or achievements, while *i*-predicates are all statives. Furthermore, since *i*-predicates are all statives, then the hypothesis that *i*-predicates are once-only predicates contradicts the characterization of statives in terms of the subset-property, as suggested for instance by Bennett and Partee (1972) and subsequent literature. Finally, Cohen and Erteschik-Shir (2002) provide the following argument against the parallelism between once-only predicates and *i*-predicates: they note that once-only predicates are fine with ‘never’, as shown in (243); but *i*-predicates are not, as shown in (244).⁸

- (243) a. John never built Mary’s house.
 b. John (has) never killed his pig.
 c. Anil will never die.
- (244) a. #John has never been Italian.
 b. #John has never been from Italy.
 c. #John has never known French.

De Swart’s proposal is very close in spirit to the one I am actually defending in this work. De Swart singles out the special class of *s*-predicates that she calls once-only predicates, and suggests that all that’s special about *i*-predicates is that they are once-only predicates. Analogously, I have singled out the class of homogeneous *s*-predicates, in the sense of (42b), and suggested that all that’s special about *i*-predicates is that they are homogeneous predicates. In this chapter, I will try to argue that my characterization of ILPs as homogeneous predicates is better off than de Swart’s characterization as once-only predicates. In particular, recall from the last remark in section 2.4, that homogeneous predicates correctly pattern with *i*-predicates in disallowing universal negative quantifiers, such as ‘never’ or ‘no’.

A missing generalization I would like to suggest that Kratzer’s original account based on (236), Jäger’s alternative account based on (239) and Swart’s account based on the analogy with only-once predicates cannot be on the right track, because they miss a crucial generalization, illustrated by the pattern in (245). Sentences (245a) and (245b) are variants of sentences (233b) and (234b), with the *i*-predicate ‘to come from a beautiful country’. According to Kratzer’s, Jäger’s and de Swart’s accounts, a sentence containing an *i*-predicate and the adverb ‘sometimes’ should turn fine as soon as the adverb is given an argument different from time to quantify over, such as the individual variable introduced by an indefinite or a bare plural. This line of reasoning incorrectly predicts the case of sentence (245c) to pattern with (245b) and thus to be fine, while instead it patterns with (245a) and sounds odd.⁹

this the NON-VACUITY PRESUPPOSITION. It is the non-vacuity presupposition that appears to be violated in the examples in [(241b)]. We can relate the deviance of [(241a)] to the deviance of [(241b)] on principled grounds. The tendential stability of *i*-predicates triggers a presupposition that there is going to be at most one state of the relevant sort, which clashes with the non-vacuity presupposition of Q-adverbs.” It seems to me non-trivial to spell out this idea on the background of Chierchia’s assumptions. Of course, only-once predicates do not admit the habitual reading, as shown in (i).

- (i) a. ? John usually kills Fido cruelly.
 b. ? John kills Fido cruelly.

Yet, Chierchia assumes that the *i*-predicate ‘tall’ in a plain sentence such as ‘John is tall’ has a time argument mandatorily bound by the same covert generic operator involved in the habitual reading of *s*-predicates.

⁸Sentences (244a) and (244b) are distinctively odd; sentence (244c) is better; and the following sentence (i) sounds perfect.

- (i) John has never been tall.

I have no proposal to make concerning this effect.

⁹Example (245c) was pointed out to me by Danny Fox (p.c.).

- (245) a. #John sometimes comes from a beautiful country.
 $\text{SOMETIMES}_t[\llbracket\text{John comes from...}\rrbracket]$
 b. Firemen sometimes come from a beautiful country.
 $\text{SOMETIMES}_x[\llbracket\text{fireman}\rrbracket(x) \wedge \llbracket\text{comes from...}\rrbracket(x)]$
 c. #Italians sometimes come from a beautiful country.
 $\text{SOMETIMES}_x[\llbracket\text{Italians}\rrbracket(x) \wedge \llbracket\text{comes from...}\rrbracket(x)]$

I will suggest in the next subsection that the correct generalization underlying the distribution of ‘sometimes’ depends on whether it quantifies over a homogeneous argument or not: sentence (245a) sounds odd because ‘sometimes’ is quantifying over time and the i-predicate ‘to come from a beautiful country’ is homogeneous with respect to that argument; sentence (245b) sounds fine because ‘sometimes’ is quantifying both over time and over firemen and the i-predicate ‘to come from a beautiful country’ is not homogeneous with respect to firemen; sentence (245c) sounds odd because ‘sometimes’ is quantifying both over time and over Italians and the i-predicate ‘to come from a beautiful country’ is homogeneous with respect to both arguments.

3.1.2 An account based on blind and mandatory mismatching implicatures

In this subsection, I introduce certain assumptions on Q-adverbs, independently of i-predicates; and then present my account for the pattern in (233)-(234), based on the assumptions (226)-(227) that i-predicates are homogeneous s-predicates.

Preliminaries Let me present my assumptions on Q-adverbs, by discussing the syntax and the semantics of the fine sentence (233a), with the s-predicate ‘available’. I assume that the LF corresponding to this sentence is roughly (246): ‘sometimes’ has the syntax of a standard determiner, whose first argument is a predicate over times C assigned by the context.

- (246) [[Sometimes_{*t*} $C(t)$] John is available at t]

In computing the truth-conditions of LFs such as (246), I will avail myself of two assumptions that have already been defended in the literature. My *first* assumption is that a predicate such as ‘available’ or ‘tall’ (indeed, any predicate but a handful of exceptions, such as ‘famous’) presupposes that its argument x is located at every time at which the predicate holds of x ; see Musan (1995) for discussion. I illustrate this assumption in (247) for the case of ‘available’. Recall that $\text{in}^w(x, t)$ is true iff the individual x is alive at time t w.r.t. the possible world w .

- (247) $\llbracket\text{available}\rrbracket^w(t) \doteq \lambda x : \text{in}^w(x, t) . x \text{ is available at } t \text{ in } w$.

My *second* assumption is (248): following for instance Berman (1990) and Schubert and Pelletier (1989), I assume that, in the computation of the truth conditions of an LF of the form $\llbracket[\text{Adv } \alpha] \beta\rrbracket$ where ‘Adv’ is any Q-adverb such as ‘sometimes’, ‘always’, etcetera, the presupposition β_{pr_s} of the nuclear scope β gets “added” via conjunction to the restrictive clause α .¹⁰

¹⁰Various arguments have been put forward in favor of assumption (248). Here is one more argument. Building on an example discussed in Larson and Segal (1995, pp. 333-334), Kai von Stechow (MIT lecture notes, Fall 2006) notes that sentence (ia) is intuitively false in the context considered, while sentence (ib) is intuitively true.

- (i) I am in a big room with five doors, all closed. After a while, John starts to go in and out of the room using four of its doors and leaving each of the four doors he used open.
 a. Every door is open.
 b. John left every door open.

The fact that sentence (ib) sounds true is puzzling, since its natural truth conditions (iia) are false in the scenario considered.

- (ii) a. $\text{every}_x[\text{door}(x)][\text{John left } x \text{ open}]$
 b. $\text{every}_x[\text{door}(x) \wedge \text{was-open}(x)][\text{John left } x \text{ open}]$

Assumption (248) offers a natural way out: assume that the nuclear scope ‘John left x open’ presupposes that “ x was open”; by assumption (248), this presupposition needs to be added by conjunction to the restrictive clause, as in (iib); the fact that sentence (ib) feels true in the scenario considered is now predicted by the fact that the truth conditions (iib) are indeed true in that scenario.

(248) The LF $[[\text{Adv } \alpha] \beta]$ denotes truth iff Adv-many times t which satisfy both α and the pre-supposition β_{pre} , also satisfy β .

Once these two assumptions (247) and (248) are in place, the LF (246) straightforwardly yields the truth conditions in (249). These truth conditions contain an existential operator \exists_t which ranges over times; its restrictive clause is the set of times in the life span $\lambda t . \text{in}^w(j, t)$ of John which satisfy the contextually assigned predicate C^w ; its nuclear scope is the set of those times at which John is available.

(249) $\lambda w . \exists_t [\text{in}^w(j, t) \wedge C^w(t)] [[\text{available}]^w(j, t)]$.

Finally, let me turn to sentences with ‘sometimes’ and an indefinite or a bare plural subject, such as sentence (250). I assume that this sentence admits three different readings. In the reading (250a), ‘sometimes’ only binds the individual variable, while the time variable is bound by a covert generic operator; in the reading (250b), ‘sometimes’ only binds the time variable, while the individual variable is bound again by a covert generic operator; in the reading (250c), ‘sometimes’ acts like an UNSELECTIVE BINDER à la Lewis (1975) and binds both the individual and the time variable.

(250) Firemen sometimes drink beer.

- a. = There are some firemen who have the habit of drinking beer.
= $\lambda w . \exists_x [[\text{fireman}](x) \wedge C^w(x)] [\text{GEN}_t [C^w(t)] [[\text{drink}]^w(x, t)]]$.
- b. = Firemen in general occasionally drink beer.
= $\lambda w . \text{GEN}_x [[\text{fireman}](x) \wedge C^w(x)] [\exists_t [C^w(t)] [[\text{drink}]^w(x, t)]]$.
- c. = There are some firemen that occasionally drink beer.
= $\lambda w . \exists_{x,t} [[\text{fireman}](x) \wedge C^w(x) \wedge C^w(t)] [[\text{drink}]^w(x, t)]$.

I am now ready to turn to the case of sentences (233b) and (234b) with the i-predicate ‘tall’ and the existential adverb ‘sometimes’.

Account By assumption (86), the odd sentence (233b) comes with a matrix exhaustivity operator, as in (251). This exhaustivity operator is restricted by the set of alternatives Ω .

(251) $[\text{EXH}_\Omega [\varphi \text{ Sometimes John is tall }]]$

Under the assumption that there is no difference between s- and i-predicates, the same LF and truth conditions as (246) and (249) should be available also for the prejacent φ of the exhaustivity operator in (251), with the i-predicate ‘tall’. I have written them down in (252), where C is a contextually assigned set of times.

(252) a. $[\varphi [\text{Sometimes}_t C(t)] \text{ John is tall at } t]$
b. $\varphi = \lambda w . \exists_t [\text{in}^w(j, t) \wedge C^w(t)] [[\text{tall}]^w(j, t)]$.

Consider the LF ψ in (253a), which is the same as the LF φ in (252a), only with the existential adverb replaced by a universal adverb. Given the two assumptions (247) and (248), this LF (253a) gets the truth conditions (253b).

(253) a. $[\psi [\text{always } C] \text{ John is tall at } t]$
b. $\psi = \lambda w . \forall_t [\text{in}^w(j, t) \wedge C^w(t)] [[\text{tall}]^w(j, t)]$.

Let me assume that ‘sometimes’ and ‘always’ are Horn-mates, as stated in (254). Evidence for this assumption (254) comes from the fact that a sentence such as ‘Sometimes, my daughter cleans up the table’ suggests that she does not always take care of that. Thus, ψ in (253) is a scalar alternative of φ in (252) according to the standard definition (74) of the set of scalar alternatives $\text{Alt}(\varphi)$.

(254) $\langle \exists_t, \forall_t \rangle$ is a Horn-scale.

Of course, ψ logically entails φ .¹¹ Furthermore, φ does not logically entail ψ : just consider a world (not compatible with common knowledge \mathcal{W}_{ck}) where John is tall sometimes in C^w but not

¹¹Modulo the caveat in footnote 11, of course.

throughout C^w and note that φ is true but ψ false in such a world. Thus, ψ logically asymmetrically entails φ and is thus excludable given φ according to definition (106), as stated in (255).¹²

$$(255) \quad \mathcal{E}xcl(\varphi) = \{\psi\}.$$

On the other hand, the alternative ψ and the prejacent φ are contextually equivalent by virtue of the characterization (226)-(227) of the i-predicateon ‘tall’, as already noted in (232).

$$(256) \quad \varphi \leftrightarrow_{w_{ck}} \psi.$$

The set of alternatives Ω of the exhaustivity operator is subject to the two conditions (151a) and (151b). Since $\psi \in \mathcal{E}xcl(\varphi)$, then ψ is allowed to belong to Ω , by assumption (151a). Furthermore, since $\varphi \leftrightarrow_{w_{ck}} \psi$, then ψ must belong to Ω , by assumption (151b). In conclusion, the odd sentence (233b) denotes the proposition $EXH_{\Omega}(\varphi)$ in (257) that “John is sometimes tall but not always.” Of course, this proposition is a contradiction given common knowledge and the oddness of sentence (233b) with the Q-adverb ‘sometimes’ and the i-predicate ‘tall’ thus follows from the MH (97).¹³

$$(257) \quad EXH_{\Omega}(\varphi) = \varphi \wedge \neg\psi$$

Before I move on, let me point out the important role played by the two assumptions (247) and (248) in the account I have just presented. Suppose that I did not avail myself of these two assumptions. Then, instead of the truth conditions φ and ψ in (252b) and (253b), I would have gotten the truth conditions φ' and ψ' in (258), respectively. By reasoning in the same way as above, I would have derived the strengthened meaning $EXH_{\Omega}(\varphi') = \varphi' \wedge \neg\psi'$. The problem is that this strengthened meaning is *not* a contradiction given common knowledge for *every* possible choice of the contextually assigned predicate C : if the restrictive clause C is slightly bigger than the life span of John, then the strengthened meaning $EXH_{\Omega}(\varphi')$ is not a contradiction given common knowledge. My proposal would thus incorrectly predict sentence (233b) to be fine, but picky with respect to the values that the context can assign to the restrictor of the adverb.

$$(258) \quad \begin{array}{l} \text{a. } \varphi' = \lambda w . \exists_t [C^w(t)] [[\text{tall}]^w(j, t)]. \\ \text{b. } \psi' = \lambda w . \forall_t [C^w(t)] [[\text{tall}]^w(j, t)]. \end{array}$$

Let me now make sure that the proposal just presented does not rule out sentence (234b) too, with the bare plural subject ‘firemen’. Of course, also the LF of this sentence contains a matrix exhaustivity operator. Since there are no differences between i- and s-predicates, the truth conditions of the prejacent φ of this exhaustivity operator in this case are completely analogous to one of the three truth conditions in (250). Here, it is sufficient to pick one of the three and show that everything goes fine. Let me pick for example the truth conditions φ in (259a), corresponding to truth conditions (250c). The only scalar alternative is the one obtained by replacing ‘sometimes’ by ‘always’, whose corresponding truth conditions are ψ in (259). By reasoning as above, I conclude that sentence (234b) has the strengthened meaning $EXH(\varphi) = \varphi \wedge \neg\psi$, namely that “some fireman are sometimes tall, but that it is not true that all firemen are always tall.” This strengthened meaning is formally analogous to the one derived in (257) for the case of the odd sentence (233b). Yet, this strengthened meaning does not contradict common knowledge (227), since it is true in a world where some fireman are always tall and some others are never tall, which is of course well compatible with common knowledge.¹⁴

$$(259) \quad \text{Firemen are sometimes tall.}$$

¹²Since ψ logically asymmetrically entails φ , then ψ is a scalar alternative of φ also with respect to Fox’s alternative definition (124).

¹³Of course, the problem with conjunction pointed out in subsection 2.5.3 with sentences (220) arises also here, for the case of sentences where the i-predicate is conjoined with an s-predicate, as in (i)

(i) #Sometimes, John is tall and reads the New York Times.

The tentative suggestion put forward in footnote 44 extends to this case. In any event, cases such as (i) seem to me to be problematic also for other existing accounts, such as Kratzer’s.

¹⁴Here, I have concentrated on the perilous truth conditions (259a), repeated in (ia), whereby the time argument is existentially quantified. Later on in section 3.3, I will suggest that the existential quantifier \exists_x is a scalar alternative of the free variable x . Suppose that in the case of φ in (ia), I could construct an alternative as follows: first, I replace the existential quantifier $\exists_{x,t}$ with the free variable x and then I bound the time argument t with a (properly restricted) universal operator \forall_t ,

- a. $\varphi = \exists_{x,t}[[\text{fireman}]](x) \wedge C(x) \wedge C(t)[[\text{tall}]](x, t)$.
 b. $\psi = \forall_{x,t}[[\text{fireman}]](x) \wedge C(x) \wedge C(t)[[\text{tall}]](x, t)$.

Finally, let me discuss the case of sentence (245c). This sentence was found above to be problematic for existing accounts for the distribution of Q-adverbs, since it has the same structure as sentence (234b), and yet the former sounds odd while the latter sounds fine. On the contrary, my account is able to account for this difference. The truth conditions corresponding to sentence (245c) are φ in (260a), that are of course completely analogous to those in (259a) for the case of sentence (234b). The truth conditions of the corresponding alternative with ‘sometimes’ replaced by ‘always’ are ψ in (260b). By reasoning as above, I conclude once more that sentence (245c) has the strengthened meaning $\text{EXH}(\varphi) = \varphi \wedge \neg\psi$. The crucial point is that this strengthened meaning does indeed contradict common knowledge. In fact, consider a world w compatible with common knowledge where the prejacent is true. This means that there exists an Italian and a time such that that Italian comes from a beautiful country at that time. By virtue of the pieces of common knowledge that Italians come from the same country and that to come from a given country satisfies the homogeneity assumption (227), it follows that every Italian satisfies the predicate at every time t in his life span in this world w , namely that the alternative ψ is true too.

(260) Italians sometimes come from a beautiful country.

- a. $\varphi = \exists_{x,t}[[\text{italians}]](x) \wedge C(x) \wedge C(t) \wedge \text{in}(x, t)[[\text{come from a beautiful country}]](x, t)$
 b. $\psi = \forall_{x,t}[[\text{italians}]](x) \wedge C(x) \wedge C(t) \wedge \text{in}(x, t)[[\text{come from a beautiful country}]](x, t)$

In conclusion, the right generalization about the distribution of ‘sometimes’ seems to be that it is only fine when it quantifies over some non-homogeneous argument. My proposal is well suited to account for this generalization. In the rest of this section, I qualify my proposal with a few remarks.

First remark: interval statives A rather interesting case is that of predicates such as ‘sit’, ‘stand’ and ‘flow’, which Dowty (1979, pp. 173-174) calls *interval statives*. Note the following contrast:¹⁵

- (261) a. Sometimes, John sits on that couch.
 b. #Sometimes, New Orleans sits on the Mississippi river.
 (262) a. Your glass sometimes sits near the edge of the table.
 b. #John’s house sometimes sits at the top of the hill.
 (263) a. The long box sometimes stands next to the door.
 b. #The new building sometimes stands at the corner of first Avenue and Main Street.

thus getting the alternative ψ'_x in (ib). Then, truth conditions (ia) would correspond to the strengthened meaning $\varphi \wedge \bigwedge_x \neg\psi_x$ that is indeed a contradiction given common knowledge. Here, I am assuming that ψ_x in (ib) is not a scalar alternative of φ in (ia), namely that the existential quantifier $\exists_{x,t}$ denoted by unselective ‘sometimes’ can only be replaced by the universal quantifier $\forall_{x,t}$ denoted by unselective ‘always’.

- (i) a. $\varphi = \exists_{x,t}[[\text{fireman}]](x) \wedge C(x) \wedge C(t)[[\text{tall}]](x, t)$.
 b. $\psi'_x = \forall_t[[\text{tall}]](x, t)$.

In any event, in order to account for the felicity of sentence (234b), it is enough to exhibit one LF which is fine. Thus, consider the LF where the prejacent φ of the matrix exhaustivity operator of sentence (234b) gets the truth conditions φ in (ii), analogous to the truth conditions in (250a).

- (ii) $\varphi = \lambda w. \exists_x [[\text{fireman}]](x) \wedge C^w(x) [\text{GEN}_t [\text{in}^w(t)] [[\text{tall}]^w(x, t)]]$.

These truth conditions contain a wide scope existential quantifier and their strengthened meaning is therefore vacuous; cf. the First Remark in section 3.3. Thus, I could have also accounted for the felicity of sentence (234b) by using these alternative truth conditions.

¹⁵Dowty makes the same point with the contrast in (i), concerning the progressive.

- (i) a. The socks are lying under the bed.
 b. #New Orleans is lying at the mouth of the Mississippi river.

In this work, I do not consider the case of the progressive; thus, I have rephrased Dowty’s observation for the case of existential Q-adverbs.

- (264) a. One leg of the piano sometimes rests on the carpet.
 b. #That argument sometimes rests on an invalid assumption.

As Fernald (2000, p. 7) puts it, these predicates “are taken to be i-predicates just in case the subject denotes the sort of thing that cannot ordinarily change its physical position from the one described by the predicate (New Orleans always lies at the mouth of the Mississippi river).” I think that the contrasts in (261)-(264) highlight the general fact that the i- vs. s-level distinction does not really apply at the level of *predicates* but rather at the level of (atomic) *sentences*, since it depends also on the arguments of the predicate. Thus, theories which encode the distinction by positing special grammatical properties for i-predicates, such as Kratzer’s assumption (236), are ill-suited to account for cases such as (261)-(264). My assumption (227) is potentially better suited, given that it is not really an assumption on the predicate $\llbracket \text{tall} \rrbracket$ but rather an assumption on the proposition $\llbracket \text{tall} \rrbracket(\text{John})$. In other words, it is very well compatible with the refinement that $\lambda t . \llbracket \text{VP} \rrbracket(\text{New Orleans}, t)$ is a homogeneous set of times for some predicates VP, while $\lambda t . \llbracket \text{VP} \rrbracket(\text{John}, t)$ is not.

Second remark: truth conditions of a plain sentence with an i-predicate Consider a plain sentence containing an i-predicate, such as (265a). According to assumption (226), the i-predicate ‘tall’ has a time argument, just as s-predicates. This argument needs to be bound at LF, in order for the sentence to denote a truth value.

- (265) a. John is tall.
 b. $[\text{GEN}_t [\text{C}(t) [\text{John tall}(t)]]]$.
 c. $\lambda w . \text{GEN}_t [\text{C}^w(t) \wedge \text{in}^w(j, t)] [\llbracket \text{tall} \rrbracket(j, t)]$.

I assume that Natural Language has two covert tense operators: an existential quantifier \exists_t , whose meaning is akin to that of overt ‘sometimes’; and a universal quantifier GEN_t called the GENERIC OPERATOR, whose semantics is akin to that of overt ‘usually’ or ‘always’. In the case of s-predicates, the time argument can be bound by either operator, yielding the episodic and the habitual reading, respectively. Since there are no grammatical differences between s- and i-predicates, then both options are in principle available also for i-predicates. Yet, we just saw in this section that existential quantification over the time argument of an i-predicate is ruled out by the MH, since it triggers the blind mismatching implicature that the corresponding sentence with universal quantification is false. Hence, the time argument of the i-predicate in (265a) can only be bound by the generic operator, namely the sentence only admits the LF in (265b). The restrictive clause of the generic operator is an arbitrary predicate of times C^w provided by context. Furthermore, by assumptions (247)-(248), the presupposition $\text{in}^w(j, t)$ that John is alive gets added to this restrictive clause, thus effectively yielding the truth conditions (265c). In the next section, I will argue that the only available choice for this predicate C is the one whereby $\text{C}^w(t) \wedge \text{in}^w(j, t) = \text{in}^w(j, t)$.

3.2 Temporal modification

As shown in (266a), temporal modification is banned with i-predicates. Furthermore, Kratzer (1995) notes that temporal modification through past-tense morphology in (266b) yields the inference that John has got to be dead. Musan (1997) calls this inference the LIFE-TIME EFFECT.

- (266) a. #John is tall after dinner.
 b. John was tall. \rightsquigarrow John is dead

In this section, I argue that the two facts in (266) correspond to the two columns of the generalization (142) discussed in section 2.3.

3.2.1 Existing accounts

According to Kratzer’s (1995) assumption (236), sentence (266a) is ruled out by the fact that the i-predicate ‘tall’ introduces no time argument for ‘after dinner’ to modify, as in (267a). Furthermore, Kratzer assumes that past morphology denotes the property of ‘belonging to the past’ and

that this property applies to both times and individuals. She thus posits the truth conditions (267b) for sentence (266b), whereby the lifetime effect follows as an entailment. As far as I can see, this same line of analysis can be repeated once Kratzer's original strong assumption (236) is replaced by the weaker version (239). I have no specific arguments against this line of account for the facts in (266), but dismiss it based on my discussion of the two crucial assumptions (236) and (239) in the preceding subsection 3.1.1.¹⁶

- (267) a. $\lambda t. \llbracket \text{after-dinner} \rrbracket(t) \wedge \llbracket \text{tall} \rrbracket(j)$
 b. $\text{PAST}(j) \wedge \llbracket \text{tall} \rrbracket(j)$.

Maienborn (2004) suggests a pragmatic account for the oddness of sentences like (266a);¹⁷ and Musan (1997) suggests a pragmatic account for the life-time effect displayed by sentences such as (266b). Their accounts are very close in spirit to the account that I will suggest in this section. Let me thus review their specific formalization as a way of motivating my own formalization. I will concentrate on Musan's proposal; Maienborn's proposal is less explicit, but it seems to me substantially analogous to Musan's proposal. Musan (1997, (22)) accounts for the life-time effect triggered by i-predicates in the past tense by means of the piece of reasoning in (268).

- (268) a. "The speaker has expressed the proposition [(269b)]."
 b. "Thus, the speaker is maximally informed about Gregory's being from America — in particular about the duration of Gregory's being from America."
 c. "If the speaker thought that Gregory's being from America is not over, he would have expressed the proposition (269a), since that would have been a more informative alternative utterance about the duration of Gregory's being from America."
 d. "Thus, the speaker couldn't have been maximally informative about Gregory's being from America unless he thought that Gregory's being from America is over."
 e. "Thus, the speaker has implicated that Gregory's being from America is over."
 f. "Since being from America is a property that, if it holds of an individual at all, holds of that individual over its entire lifespan, and since the speaker has implicated that Gregory's being from America is over, the speaker has implicated furthermore that Gregory is dead."
- (269) a. Gregory is from America.
 b. Gregory was from America.

The crucial step of this reasoning is (268c), which says that (269a) is "more informative" than (269b). But I do not understand the way Musan argues for this claim.¹⁸ She assumes the semantics in (270) for the two sentences (269), where t is a time interval. Crucially, these truth conditions have existential quantification over time. I do not understand why truth-conditions (270a) are more informative than truth conditions (270b).

- (270) a. $\llbracket \text{Gregory is American} \rrbracket(\text{NOW}) = 1 \iff \exists t[\text{NOW} \in t \wedge \text{AMERICAN}(g, t)]$
 b. $\llbracket \text{Gregory was American} \rrbracket(\text{NOW}) = 1 \iff \exists t[t < \text{NOW} \wedge \text{AMERICAN}(g, t)]$

¹⁶Musan's (1997) arguments against Kratzer's analysis only target her further assumption that past morphology can only apply to the external argument, either the Davidsonian argument (for s-predicates) or the subject (for i-predicates).

¹⁷She considers sentences containing locatives rather than tense modifiers, such as 'Mary is blond in her car', but her proposal seems to me to straightforwardly extend to tense modifiers too.

¹⁸Musan explains this claim in the following passage, that I do not understand: "Suppose (269a) is true. In this case, we know the following: if (269b) is also true, then the situation time of 'be from America' obviously reaches into the past (because of the truth of (269a)), i.e., the implication from the present tense clause to the past tense clause is guaranteed. But how about the case where (269a) is false? For practical purposes in a concrete discourse, this possibility can be disregarded because conversation takes place under the assumption that utterances are truthful. Hence, when [(269a)] is uttered, for practical purposes — which only care about cases where the past tense clause is true — the present tense clause is justified to count as more informative than the past tense clause. It seems that this relationship justifies treating past tense clauses and present tense clauses as ordered with regard to informativeness" (pp. 280-281).

The problem here is that there is no subset relationship between the two sets $\llbracket \text{PAST} \rrbracket = \lambda t . t < \text{NOW}$ and $\llbracket \text{PRES} \rrbracket = \lambda t . \text{NOW} \in t$, and thus no way to compare the informativeness of the two propositions (269). In order to get a subset relationship, I will follow below Sauerland (2002) and assume that the present tense is vacuous, namely that it does not impose any restriction. This way, we do get a subset relationship $\llbracket \text{PAST} \rrbracket \subseteq \llbracket \text{PRES} \rrbracket$. Yet, if we stick with Musan’s assumption (270) that time is existentially quantified, then we make (269b) stronger and thus more informative than (269a), contrary to what we want. To get out of trouble, I will suggest that time in these two sentences is not quantified by an existential operator but rather by a universal operator and that tense morphemes end up in the restrictor of that universal operator, which is a DE environment.

3.2.2 An account based on blind and mandatory mismatching implicatures

In this subsection, I present what I take to be a streamlined version of Musan’s and Maienborn’s accounts for the facts in (266).

Preliminaries Let me start with a naïve way of restating Musan’s and Maienborn’s proposals, in order to see where it fails. Building on the last Remark in subsection 3.1.2, I assume that the truth conditions of sentence (266a) are plausibly φ_1 in (271a), where the temporal modifier ‘after dinner’ further restricts the restrictive clause of the generic operator. Furthermore, I assume that the truth conditions of sentence (266b) are φ_2 in (271b), namely that past-tense morphology acts as a tense modifier, analogously to ‘after dinner’.

- (271) a. $\varphi_1 = \lambda w . \text{GEN}_t [\text{in}^w(j, t) \wedge \llbracket \text{after-dinner} \rrbracket^w(t)] \llbracket \llbracket \text{tall} \rrbracket^w(j, t) \rrbracket$.
 b. $\varphi_2 = \lambda w . \text{GEN}_t [\text{in}^w(j, t) \wedge \llbracket \text{PAST} \rrbracket^w(t)] \llbracket \llbracket \text{tall} \rrbracket^w(j, t) \rrbracket$.

Assume that modifiers in the restrictor of a universal operator can be dropped without falling out of the set of scalar alternatives. Support for this assumption comes from the fact that the sentence ‘Every *blond* woman is beautiful’ triggers the implicature that not every woman is beautiful. Hence, ψ in (272) is a scalar alternative of φ_1 in (271a). Assume furthermore that present- and past-tense morphology are Horn-mates and that present-tense morphology is semantically vacuous, as suggested for instance in Sauerland (2002); hence, ψ in (272) is a scalar alternative of φ_2 in (271b) too.

- (272) $\psi = \lambda w . \text{GEN}_t [\text{in}^w(j, t)] \llbracket \llbracket \text{tall} \rrbracket^w(j, t) \rrbracket$.

Of course, ψ in (272) logically asymmetrically entails both φ_1 and φ_2 in (271), since the latter two formulas have some extra stuff in their restrictors. Hence, the blind strengthened meaning of both φ_1 and φ_2 in (271) is predicted to be the one in (273).

- (273) $\text{EXH}(\varphi_i) = \varphi_i \wedge \neg\psi$, for $i = 1, 2$.

In the case of sentence (266a) with truth conditions φ_1 , this blind strengthened meaning says that “John is always tall after dinner but there are non-after-dinner times at which he is not tall”; of course, this blind strengthened meaning contradicts the piece of common knowledge (227) that tallness is a permanent property; in conclusion, sentence (266a) is correctly predicted to sound odd. The problem is that exactly the same prediction is made for the case of sentence (266b) with truth conditions φ_2 : its blind strengthened meaning says that “John was always tall in the past but there are times not in the past where John is not tall”; again, this blind strengthened meaning contradicts the common knowledge that tallness is a permanent property; thus, sentence (266b) is incorrectly predicted to sound odd too and furthermore no life-time effect is derived. What is the relevant difference between the two sentences (266)? I would like to argue that the discussion presented in Subsection 2.3 sheds some light on this puzzle.

Account The relevant scalar items in the case of the two sentences (266) are the tense modifier ‘after dinner’ and the past-tense morphology ‘PAST’. According to the truth conditions (271), these scalar items are embedded in a DE environment, namely the restrictor of a universal quantifier. Thus, the generalization in table (142) applies; let’s see what it predicts. Let’s start with the case of sentence (266a), with the truth conditions φ_1 in (271a) and the alternative ψ in (272). They correspond to the

target sentence $\varphi = \text{every}(\text{Strong})$ and the alternative $\psi = \text{every}(\text{Weak})$ in table (142) with the two restrictors *Strong* and *Weak* defined as in (274). Note that this is a case where the two restrictors cannot be equivalent given common knowledge: in fact, the equivalence $\text{Strong} \leftrightarrow_{\mathcal{W}_{\text{ck}}} \text{Weak}$ would mean that it follows from common knowledge \mathcal{W}_{ck} that John is only alive at after dinner times, which obviously cannot be. Hence, the right column of table (142) applies in this case: sentence (266a) with truth conditions φ_1 with the stronger restrictor is correctly predicted to sound odd; sentence ψ with the weaker restrictor *Weak* is predicted to sound fine.

- (274) a. $\text{Strong}^w = \lambda t . \text{in}^w(j, t) \wedge [\text{after-dinner}]^w(t)$.
 b. $\text{Weak}^w = \lambda t . \text{in}^w(j, t)$.

Let's now turn to the case of sentence (266b), with the truth conditions φ_2 in (271b) and the alternative ψ in (272). Again, they correspond to the target sentence $\varphi = \text{every}(\text{Strong})$ and the alternative $\psi = \text{every}(\text{Weak})$ in table (142) with the two restrictors *Strong* and *Weak* defined as in (275). Two cases need to be considered. One case is that it follows from common knowledge that John is already dead. Equivalently, that two restrictors *Strong* and *Weak* in (275) are equivalent given common knowledge. In this case, the left column of table (142) applies: sentence (266b) with truth conditions φ_2 with the past tense predicate is predicted to sound fine and the alternative ψ without the past tense predicate is predicted to sound odd. The other case that needs to be considered is that it does not follow from common knowledge that John is already dead. Equivalently, that the two restrictors *Strong* and *Weak* in (275) are *not* equivalent given common knowledge. In this case, the right column of table (142) applies: sentence (266b) with truth conditions φ_2 with the past tense predicate is predicted to sound odd and the alternative ψ without the past tense predicate is predicted to sound fine. The life-time effect is thus derived.¹⁹

- (275) a. $\text{Strong}^w = \lambda t . \text{in}^w(j, t) \wedge [\text{PAST}]^w(t)$.
 b. $\text{Weak}^w = \lambda t . \text{in}^w(j, t)$.

In conclusion, the two tense modifiers 'after dinner' and PAST in sentences (266) lead to two different effects, namely oddness and the life-time effect, respectively; these two effects correspond to the two columns of table (142). The proposal made in section 2.3 to account for the generalization (142) thus extends to the case of the two sentences (266). For the sake of explicitness, let me quickly recall the basic idea of that account. The LFs of both sentences (266) contain an instance of the exhaustivity operator embedded in the restrictor of the generic operator. In the case of sentence (266a), the two restrictors *Strong* and *Weak* in (274) are not contextually equivalent, hence the embedded implicature is not mandatory and thus the embedded exhaustivity operator ends up being vacuous. In the case of sentence (266b) where common knowledge entails that John is dead, the two restrictors *Strong* and *Weak* in (275) are contextually equivalent, thus the embedded implicature is mandatory and it leads to a mismatch.

First remark: more on the truth conditions of a plain sentence with an i-predicate Consider a plain sentence containing an i-predicate, such as (276a). As discussed in the Second Remark in subsection 3.1.2, the time argument of the i-predicate 'tall' in sentence (276a) can only be bound by the generic operator, yielding the truth conditions (265c), repeated in (276b), whereby the generic operator is restricted by the predicate of times $\lambda t . \text{C}^w(t) \wedge \text{in}^w(j, t)$, where *C* is contextually supplied. Yet, as seen in this section, any choice of the predicate *C* such that there are worlds compatible with common knowledge where $\text{C}^w(t) \wedge \text{in}^w(j, t) \neq \text{in}^w(j, t)$ is ruled out by the fact that it leads to a blind mismatching implicature. In conclusion, the only truth conditions predicted for sentence

¹⁹Consider sentence (i), with the present tense morphology. What I have derived is not that the felicity of sentence (i) requires common knowledge to entail that John is alive but rather something weaker, namely that common knowledge does not entail that he is dead.

- (i) John is tall.
 (ii) A: Do you know John? do you know if he still alive?
 B: That I don't know. I know he is tall and has blue eyes.

I think that this prediction might not be wrong, since the dialogue in (ii) might be fine.

(276a) are those in (276c), at least if we restrict ourselves to possible worlds compatible with common knowledge.

- (276) a. John is tall.
 b. $\lambda w . \text{GEN}_t[\text{in}^w(j, t) \wedge \mathbf{C}^w(t)] [[\text{tall}]](j, t)$.
 c. $\lambda w . \text{GEN}_t[\text{in}^w(j, t)] [[\text{tall}]](j, t)$.

Truth conditions (276c) are precisely those posited by Chierchia (1995). Thus, I have derived Chierchia's assumption that (276c) are the only possible truth conditions for sentence (276a) without any *ad hoc* assumptions.²⁰

Second remark: licit tense modifiers I-predicates do not *always* ban temporal modification: although sentence (277a) with the temporal modifier 'after dinner' sounds odd, sentence (277b), with the temporal modifier 'since he was fifteen', sounds fine. This contrast is difficult to derive within a framework such as Kratzer's. Assume that the oddness of sentence (277a) is due to the fact that 'to know Latin' lacks a Davidsonian argument by (236) and thus 'after dinner' has nothing to modify. In order to reconcile this account with the felicity of (277b), we would have to say that the i-predicate has been coerced into an s-predicate in this case. But then we are unable to account for the oddness of (277c): we would have expected that also in the case of (277c), just as in the case of (277b), the i-predicate can be coerced into an s-predicate in order to provide an argument for 'since he was fifteen' to modify, offering a free ride to the further modifier 'after dinner'. Clearly, something else needs to be said about this latter case (277c). Whatever we say about (277c) will plausibly apply also to (277a), so that the assumption (236) that i-predicates lack a Davidsonian argument that we started with ends up being redundant.

- (277) a. #John knows Latin after dinner.
 b. John knows Latin since he was fifteen
 c. #John knows latin after dinner since he was fifteen.

Here is how my proposal might cope with the contrast between (277a) and (277b). My current characterization (227) of i-predicates says that in every world $w \in \mathcal{W}_{\text{ck}}$ compatible with common knowledge, if an individual x satisfies an i-predicate P^w at a time t' , then x satisfies the predicate P^w at every time t at which x is alive, namely at every time t such that $\text{in}^w(x, t)$. This characterization is supposed to capture the intuition that what's special about i-predicates is that it follows from common knowledge that they are permanent, that they tend to last. As already noted at the beginning of the chapter, this current formulation (227) cashes out this intuition in a rather rough way. One might want to be more careful, and replace it with the following more fine grained variant (278).

- (278) In every world in $w \in \mathcal{W}_{\text{ck}}$ compatible with common knowledge, if an individual x satisfies an i-predicate P^w at a time t , then x satisfies the predicate *at least* at every time t such that $\mathbf{C}_P^w(x, t)$.

Let me call \mathbf{C}_P the ASSUMED MINIMAL DURATION of the i-predicate P . According to assumption (227), the assumed minimal duration $\mathbf{C}_P(\cdot, \cdot)$ is $\text{in}(\cdot, \cdot)$ for every i-predicate P . Assumption (278) improves on assumption (227) in allowing different i-predicates to have different assumed minimal durations. For instance, it is plausible that the assumed minimal duration $\mathbf{C}_{\text{French}}$ of the predicate 'to be French' is indeed in , since if John is French then plausibly he is French throughout his entire life. But in the case of the predicate 'to know Latin', this doesn't seem plausible and we may want instead its assumed minimal duration \mathbf{C}_{know} to be much shorter. Nothing changes (I think) in the account presented in this section if assumption (227) is replaced by assumption (278). And the pattern of judgments in (277) immediately follows.

²⁰Chierchia uses instead two specific assumptions. In order to force generic quantification rather than existential quantification over the time argument of the predicate, Chierchia assumes that i-predicates come with a special feature that forces local agreement with the generic operator. Furthermore, Chierchia (footnote 22) resorts to a special presupposition in order to rule out other restrictors besides in .

Third remark: generic numerical indefinites The plural indefinite ‘two firemen’ of sentence (279a) admits both the generic reading (“every two prototypical firemen love each other”) and the existential reading (“there are two firemen who love each other”). If we replace the collective i-predicate ‘to love each other’ with the inherently distributive i-predicate ‘tall’, we get a very different effect: the plural indefinite ‘two firemen’ of sentence (279b) only admits the existential reading (“there are two firemen who are tall”) but lacks the generic reading (“every two prototypical firemen are tall”, which would then boil down to “every prototypical fireman is tall”, because of the distributivity of the predicate ‘tall’). The contrast between (279a) and (279b) can be sharpened by adding an overt ‘usually’ with matrix scope (which blocks the existential reading of the indefinite) and noting that sentence (279a) remains fine while sentence (279b) turns distinctively odd.

- (279) a. Two firemen love each other. [the ‘two’-indefinite admits the generic reading]
 b. Two firemen are tall. [the ‘two’-indefinite does not admit the generic reading]

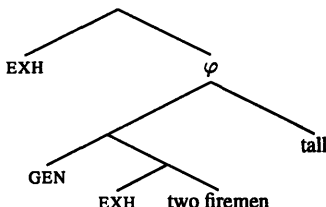
In order to account for the generic reading of the plural indefinite of sentence (279a), let me assume that ‘two firemen’ introduces a variable that ranges over plural individuals $x+y$ that are the sum of two (different) firemen x and y ; and furthermore that this variable can get bound by the generic operator, thus yielding the truth conditions (280a). Based on the preceding First Remark, I have assumed here that the restrictive clause of the generic operator is as large as possible, namely the intersection of the two life spans of the two firemen x and y that the generic operator is quantifying over. We then expect sentence (279b) to admit the truth conditions (280b), analogous to the truth conditions (280a) of sentence (279a). Here, I have assumed that the inherently distributive predicate ‘tall’ has been operated upon by a distributive operator in order to be able to apply to pairs of individuals. What is wrong with sentence (279b)?

- (280) a. $\text{GEN}_{x,y,t}[x \neq y \wedge [\text{fireman}](x) \wedge [\text{fireman}](y) \wedge \text{in}(x,t) \wedge \text{in}(y,t)] [[\text{love}]](x+y,t)$.
 b. $\text{GEN}_{x,y,t}[x \neq y \wedge [\text{fireman}](x) \wedge [\text{fireman}](y) \wedge \text{in}(x,t) \wedge \text{in}(y,t)] [[\text{tall}]](x,t) \wedge [[\text{tall}]](y,t)$.

Here is the sketch of a possible way to go. So far, I have construed the Davidsonian argument just as time. This is of course an oversimplification. So assume for instance that the Davidsonian argument is a situation, after Kratzer (2007). And let me rewrite the truth conditions (280b) as φ in (281a). Consider the scalar alternative of sentence (279b) obtained by replacing ‘two’ with ‘a’. The corresponding truth conditions are ψ (281b). Consider the proposition $\varphi \wedge \neg\psi$ that “any two firemen are tall when they are together but it is not the case that any fireman is tall at any situation where he is located”. Plausibly, this proposition is not a logical contradiction: just consider a possible world where there are only two firemen, who happen to be tall in those situations where they are together, so that φ is true; but that are not tall in those situations where they are not together, so that $\neg\psi$ is true too. Let me thus assume that the strengthened meaning of the target sentence φ is indeed $\varphi \wedge \neg\psi$.²¹

²¹Let me discuss this latter claim more carefully. In section 2.3, I have suggested that the exhaustivity operator is mandatory at every scope site, as stated in (147); and furthermore that the restrictor of universal quantifiers should be construed as an open formula, so that it requires an embedded exhaustivity operator. Thus, the LF of sentence (279b) has two instances of the exhaustivity operator, as in (i): a matrix one and another one embedded in the restrictive clause of the generic operator that bounds the indefinite and the time argument.

(i)



The only interesting alternative of the preajcent ‘two firemen’ of the embedded exhaustivity operator is the alternative obtained by replacing ‘two’ by ‘a’. The denotation of these two restrictors is given in (ii), construed as an open proposition.

- (ii) a. $[[\text{two firemen}]] = \lambda w . [[\text{two firemen}]]^w(x) \wedge [[\text{two firemen}]]^w(y) \wedge \text{in}^w(x,t) \wedge \text{in}^w(y,t)$.
 b. $[[\text{a fireman}]] = \lambda w . [[\text{fireman}]]^w(x) \wedge \text{in}^w(x,t)$.

Now we need to ask whether condition (151b) applies or not in this case, namely whether the two restrictors (ii) count as contextually equivalent or not and thus whether the embedded implicature is mandatory or not. I think the answer depends

- (281) a. $\varphi = \text{GEN}_{x,y,s}[x \neq y \wedge \llbracket \text{fireman} \rrbracket(x) \wedge \llbracket \text{fireman} \rrbracket(y) \wedge \text{in}(x,s) \wedge \text{in}(y,s)] \llbracket \llbracket \text{tall} \rrbracket(x,s) \wedge \llbracket \text{tall} \rrbracket(y,s) \rrbracket$
 b. $\psi = \text{GEN}_{x,s}[\llbracket \text{fireman} \rrbracket(x) \wedge \text{in}(x,s)] \llbracket \llbracket \text{tall} \rrbracket(x,s) \rrbracket$.

Let me assume that common knowledge entails that there are no lonely firemen, as stated in (282). Then, the proposition $\varphi \wedge \neg\psi$ is a contradiction given common knowledge and the lack of the generic reading of sentence (279b) is accounted for.

- (282) For every world $w \in \mathcal{W}_{\text{ck}}$ compatible with common knowledge and for every fireman x , there is at least a situation s and a firemen y (different from x) such that x and y are together in s .

I close this discussion with a very puzzling complication. Dobrovie-Sorin (2003) discusses plural French ‘des’-indefinites in subject position with generic meaning. She notes the following contrast: generic ‘des’-indefinites are fine with collective i-predicates, as in (283); but bad with distributive ones, as in (284).

- (283) a. Des droites convergentes ont un point en comun.
 ‘Des convergent lines have a point in common’.
 b. Des pays limitrophes ont souvent des rapports difficiles.
 ‘Des neighboring countries have frequently have difficult relations’.
 c. Des vrais jumeaux se ressemblent dans les plus petits détails.
 ‘Des true twins look alike down to the smallest details’.
- (284) a. #Des carrés ont quatre côtés.
 Des squares have four sides.
 b. #Des chats sont intelligents.
 Des cats are intelligent.

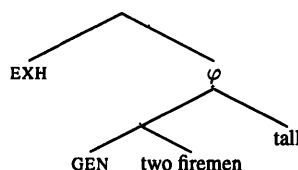
This pattern with French ‘des’-indefinites can be replicated with Italian pseudopartitive ‘dei/degli NP’; I provide an example in (285). The pattern in the case of French ‘des’-indefinites and Italian pseudo-partitives looks very similar to that in (279): as Dobrovie-Sorin observes, the pattern in (284)-(283) stays exactly the same if ‘des’ is replaced by a numeral, say ‘two’.

- (285) a. #Dei quadrati hanno quattro lati.
 of-the squares have four sides.
 b. Delle rette convergenti si incontrano in un punto.
 of-the lines convergent themselves meet in a point.

Despite this resemblance, there seems to me to be no way to extend the account sketched above for the oddness of (279b) to the case of French ‘des’-indefinites or Italian pseudo-partitives. Let me make the problem explicit. In order to extend my proposal from the former case to the latter case, I would have to assume that the property provided by ‘des’-indefinites and pseudo-partitives for the generic operator to quantify over only contains plural individuals, just as in the case of ‘two firemen’. It would then be plausible to assume that the same holds when ‘des’-indefinites and pseudo-partitives are used as existential quantifiers. Yet, this assumption would mismatch with the well known fact

on the fine grained details of the formulation of assumption (151b). If it turns out that the two restrictors (ii) do not count as contextually equivalent, then condition (151b) does not apply and the embedded exhaustivity operator can thus be effectively ignored. In other words, we can assume for all intent and purposes that the LF of sentence (279b) is (iii), with only the matrix exhaustivity operator.

(iii)



In this case, the reasoning suggested in the text goes through.

that the plurality inference of ‘des’-indefinites and pseudo-partitives (as well as any other plural existential quantifier) disappears under negation, contrary to the case of ‘two’-indefinites. This fact is illustrated in (refex: Mario nono ha comprato dei libri) for the case of Italian pseudo-partitives: sentence (286a) with the ‘two’-indefinite is true in a scenario where Mario bought only one book; sentence (286b) with the pseudo-partitive is not.

- (286) a. Mario non ha comprato due libri.
 Mario not did buy two books
 ‘Mario didn’t buy two books’
 b. Mario non ha comprato dei libri.
 Mario not did buy of-the books.
 ‘Mario didn’t buy books’

In other words, I am faced with the following conundrum: on the one hand, in order to get the contrast in (286), I need to assume that ‘des’-indefinites and pseudo-partitives are analogous to English bare plurals and different from ‘two’-indefinites, namely that ‘des’-indefinites and pseudo-partitives existentially quantify over both singular and plural individuals; on the other hand, in order to account for the analogy between (279) and (283)-(285), I need to assume that ‘des’-indefinites and pseudo-partitives are analogous to ‘two’-indefinites and different from English bare plurals, namely that ‘des’-indefinites and pseudo-partitives provide a restrictor to the generic operator that does not contain singular individuals. The mystery will get even thicker when German negative bare plurals will enter the picture in section 3.4.

3.3 Bare plural subjects

The bare plural subject (henceforth: BPS) ‘firemen’ of the s-predicate ‘available’ in (287a) admits both the generic reading (“Firemen are available people”) and the existential reading (“There are firemen who are available”); the BPS of the i-predicate ‘tall’ in (287b) instead lacks the existential reading and only allows the generic one.

- (287) a. Firemen are available. ∃-BPS, GEN-BPS
 b. Firemen are tall. #∃-BPS, GEN-BPS

In this section, I argue that the existential reading of the BPS of sentence (287b) is unattested because the blind strengthened meaning that would correspond to its LF is a contradiction given the common knowledge (227) that ‘tall’ is a permanent property. More precisely, I argue that the existential reading of the BPS of sentence (287b) can be ruled out in exactly the same way I have accounted in section 2.2 for the oddness of sentence (99a), repeated in (288a).

- (288) *Context*: a competition lasted five days, Monday through Friday; both John and Bill know that the same guy *x* won on each of the five days:

	M	T	W	T	F
<i>winner</i> :	<i>x</i>	<i>x</i>	<i>x</i>	<i>x</i>	<i>x</i>

John wants to know more about this amazing guy and thus asks Bill for more information; Bill provides the following information:

- a. #On every day, a/some fireman won.
 b. A/Some fireman won on every day.

My argument exploits the following parallelism. In the case of sentence (288a), it is crucial that the indefinite ‘a/some fireman’ has scope below the universal quantifier ‘on every day’, since sentence (288b) where the scope is reversed is acceptable. Analogously, Carlson (1977) points out that existential BPs always take the narrowest possible scope with respect to other scope bearing elements in the sentence. Furthermore, the context described in (288) entails that, if a fireman won on one day, then he won on every day; analogously, the common knowledge \mathcal{W}_{ck} described in (227) entails that, if a fireman is tall at a given time, then he always is. To stress the analogy between the account for

the oddness of sentence (288a) suggested in section 2.2 and the account for the lack of the existential reading of the BPS of sentence (287b) suggested in this section, Table 3.4 summarizes the main steps of the two accounts, one next to the other.

3.3.1 Existing accounts

Two recent approaches to the lack of the existential reading of BPSs of *i*-predicates have been developed in the literature. One approach is based on syntactic assumptions, and has been developed mainly by Kratzer (1995), Diesing (1992) and Chierchia (1995). Another approach is based on information structure assumptions, and has been developed mainly by Jäger (2001) and Cohen and Erteschik-Shir (2002). In this subsection, I review both approaches in turn and try to argue that they are not fully satisfactory, thus making room for my own proposal. An older approach due to Carlson (1977) will not be reviewed here; see for instance Wilkinson (1991) for discussion.

The syntactic approach: overall idea Diesing's and Chierchia's proposals differ substantially with respect to both their syntactic assumptions (e.g. the position at LF of subjects of *i*- and *s*-predicates; the characterization of the restrictive clause and the nuclear scope of Q-adverbs; etcetera) and their semantic assumptions (e.g. the semantics of bare plurals and indefinites; the semantics of LFs containing Q-adverbs; etcetera). Yet, the two accounts share a common logic, that is worth sketching at the beginning, before turning to the details. Both authors assume that the generic reading of BPSs comes about through a covert Q-adverb with generic force, called the **GENERIC OPERATOR** and notated **GEN**. Like any other Q-adverb, this generic operator has a restrictor and a nuclear scope. If a BP ends up in the restrictor of the generic operator, then the BP gets bound by the generic operator and receives the generic reading. If instead the BP ends up in the nuclear scope of the generic operator, then the generic operator cannot bind the BP, and the BP gets closed off existentially. Thus, the first step of Diesing's and Chierchia's accounts is (289): whether a BPS is interpreted existentially or generically depends on whether it sits at LF within or outside of the nuclear scope $[_{NS} \dots]$ of the generic operator. Diesing calls this assumption (289) a **MAPPING HYPOTHESIS**, since it constraints the mapping between LFs and truth conditions.

(289) *First step:*

- a. The BPS in the LF $[\dots [_{NS} \dots \text{BPS} \dots]]$ has only the existential reading;
- b. The BPS in the LF $[\dots \text{BPS} \dots [_{NS} \dots]]$ has only the generic reading.

The second step of Diesing's and Chierchia's account is (290), namely a constraint on the LF position of the subject of an *i*-predicate such as 'tall' with respect to the nuclear scope $[_{NS} \dots]$ of the generic operator.

(290) *Second step:*

- a. $[\dots \text{SUBJ} \dots [_{NS} \dots \text{tall} \dots]]$;
- b. $*[\dots [_{NS} \dots \text{SUBJ} \dots \text{tall} \dots]]$.

The lack of the existential reading of the BPS 'firemen' of the *i*-predicate 'tall' of sentence (287b) follows straightforwardly from these two assumptions (289) and (290): by (290), the BPS must sit outside of the nuclear scope $[_{NS} \dots]$ of the generic operator; by (289), a BPS in that LF position cannot have the existential reading. The availability of the existential reading of the BPS 'firemen' of the *s*-predicate 'available' of sentence (287a) follows just as straightforwardly, since constraint (290) does not apply to subjects of *s*-predicates so that the BPS is allowed to sit within the nuclear scope $[_{NS} \dots]$ of the generic operator in the case of 'available'. In conclusion, the two steps (289) and (290) yield an account for the contrast in (287). As it is clear even from this cursory initial description, this account is *syntactic* in nature, since it crucially relies on the assumption that the LF (290b) is ungrammatical. Diesing is particularly explicit about the syntactic nature of her account: "I show that this contrast [(287)] between the two types of predicates is actually syntactic in nature, but because of the working of the Mapping Hypothesis [(289)], it is reflected also in the available semantic interpretations of a BPS" (p. 11). Let me now discuss Diesing's implementation of these

#Each day, a/some fireman won.	#Firemen are tall. (with \exists -BPS reading)																								
(1)	LOGICAL FORM																								
<p>Tree structure: EXH branches to 'each' and φ. φ branches to 'day' and $[\text{some fireman}]_x$. $[\text{some fireman}]_x$ branches to 'won'.</p>	<p>Tree structure: EXH branches to GEN and φ. GEN branches to 'firemen_x'. φ branches to C and VP. C branches to 'firemen_x'. VP branches to 'tall'.</p>																								
(2)	TRUTH CONDITIONS OF THE PREJACENT																								
$\varphi = \lambda w . \forall_t [[\text{day}]^w(t)] [\exists_x [[\text{win}]^w(x, t)]]$	$\varphi = \lambda w . \text{GEN}_t [\tilde{\text{C}}^w(t)] [\exists_x [[\text{tall}]^w(x, t)]]$																								
(3)	ASSUMPTION ON HORN SCALES																								
⟨'some fireman', 'the fireman such and such'⟩	⟨'firemen', 'the fireman such and such'⟩																								
Notation: 'the fireman P ' stands for 'the fireman such and such' and d_P^w stands for its denotation																									
(4)	SET OF SCALAR ALTERNATIVES																								
$\text{Alt}(\varphi) = \{ \psi_P \mid P \in \mathcal{D}_{\langle e, t \rangle}^w \}$, where ψ_P is the LF with the following shape and truth conditions																									
<p>Tree structure: ψ_P branches to 'each' and 'day'. 'day' branches to $[\text{the fireman } P]$. $[\text{the fireman } P]$ branches to 'won'.</p>	<p>Tree structure: ψ_P branches to GEN and C. GEN branches to 'firemen'. C branches to $[\text{the fireman } P]$. $[\text{the fireman } P]$ branches to 'tall'.</p>																								
$\psi_P = \lambda w . \forall_t [[\text{day}]^w(t)] [[\text{win}]^w(d_P^w, t)]$	$\psi_P = \lambda w . \text{GEN}_t [\text{C}^w(t) \wedge \text{in}^w(d_P^w, t)] [[\text{tall}]^w(d_P^w, t)]$																								
(5)	A USEFUL RESTATEMENT OF THE NEGATION OF ALL THE ALTERNATIVES																								
$\bigwedge \neg \psi_P = \neg \psi$, where ψ are the following truth conditions																									
$\psi \doteq \lambda w . \exists_x \forall_t [[\text{day}]^w(t)] [[\text{win}]^w(x, t)]$	$\psi \doteq \lambda w . \exists_x [\text{GEN}_t [\text{C}^w(t) \wedge \text{in}^w(x, t)] [[\text{tall}]^w(x, t)]]$																								
(6)	SET OF EXCLUDABLE ALTERNATIVES																								
$\text{Excl}(\varphi) = \{ \psi_P \mid P \in \mathcal{D}_{\langle e, t \rangle}^w \}$, since $\varphi \wedge \neg \psi$ is true e.g. in the following world:																									
<table border="1" style="border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th>M</th> <th>T</th> <th>W</th> </tr> </thead> <tbody> <tr> <td>John:</td> <td></td> <td></td> <td>✓</td> </tr> <tr> <td>Bill:</td> <td></td> <td>✓</td> <td></td> </tr> <tr> <td>Tom:</td> <td>✓</td> <td></td> <td></td> </tr> </tbody> </table>		M	T	W	John:			✓	Bill:		✓		Tom:	✓			<table style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th></th> <th style="text-align: center;">$\tilde{\text{C}}^w(t)$</th> </tr> </thead> <tbody> <tr> <td>John:</td> <td style="text-align: center;">—————</td> </tr> <tr> <td>Bill:</td> <td style="text-align: center;">—————</td> </tr> <tr> <td>Tom:</td> <td style="text-align: center;">—————</td> </tr> </tbody> </table>		$\tilde{\text{C}}^w(t)$	John:	—————	Bill:	—————	Tom:	—————
	M	T	W																						
John:			✓																						
Bill:		✓																							
Tom:	✓																								
	$\tilde{\text{C}}^w(t)$																								
John:	—————																								
Bill:	—————																								
Tom:	—————																								
(7)	CONCLUSION																								
the sentences denote the proposition $\text{EXH}(\varphi) = \varphi \wedge \bigwedge_P \neg \psi_P = \varphi \wedge \neg \psi$ that contradicts common knowledge; oddness thus follows from the MH (97)																									

Table 3.4: Parallelism between the account for the oddness of sentence (288a) presented in section 2.2 and the account for the oddness for sentence (233b) presented in this section.

two steps (289) and (290); similar considerations apply to Chierchia's implementation, that I will not discuss in detail here.

Diesing's version of the syntactic approach Building on Heim (1988) and Kamp (1981), Diesing posits the three assumptions in (291). According to these assumptions, the nuclear scope of the generic operator is VP; BPSs within VP are existentially closed off by the DEO; BPSs outside of VP are bound by the generic operator. The first step (289) of the syntactic approach, concerning the correlation between LF position of a BPS and the availability of its existential or generic construal, thus immediately follows from these assumptions (291).

- (291) a. A BP $[_{DP} \emptyset N]$ has no quantificational force on its own: it simply introduces a free variable x which ranges over $[N]$.
 b. There is a covert default existential operator (henceforth: DEO) that binds every free variable in its scope; its scope is defined to be VP.
 c. A Q-adverb ('always', 'sometimes', GEN, etcetera) binds every free variable outside of its nuclear scope; its nuclear scope is defined to be VP.

On the background of these assumptions (291), the second step (290) reduces to the assumption that subjects of i-predicates must be outside of VP at LF. Two different implementations of this assumption have been suggested in the literature: the one in (292) is due to Kratzer (1995); the one in (293) is due to Diesing (1992). The advantage of Kratzer's version (292) is that it can be linked to her assumption (236) that i-predicates lack a Davidsonian argument, as follows. The "external argument" of a predicate is the argument that is base generated outside of the maximal projection of the predicate. Each predicate comes with one and only one external argument. S-predicates have a Davidsonian argument that counts as their external argument; hence, their subjects must be VP internal. I-predicates lack the Davidsonian argument by assumption (236) and thus it is their subjects that count as their external argument.²² Diesing replaces (292) with (293) based on the observation that floated quantifiers are fine with i-predicates and plausibly require the [Spec, VP] position to be filled.²³

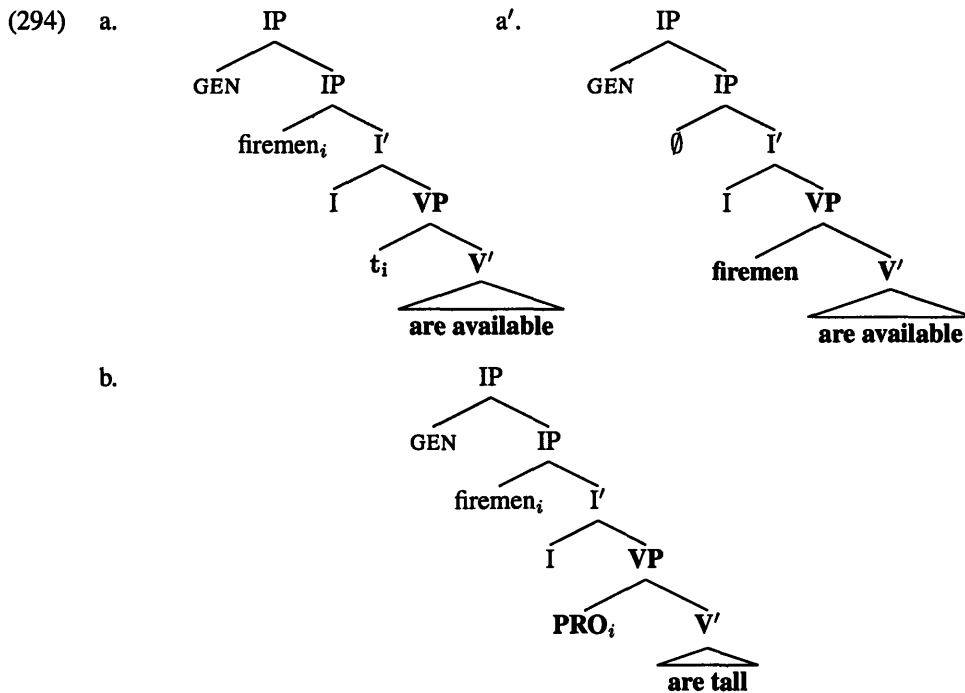
- (292) a. Subjects of s-predicates are base generated in [Spec, VP] and can thus be reconstructed into VP at LF;
 b. subjects of i-predicates are based generated in [Spec, IP] while [Spec, VP] is left empty, hence they cannot be reconstructed into VP at LF.
- (293) "The difference between the two types of predicates [i-/s-predicates] arises from differences in the properties of the Infl associated with them" (p. 24):
 a. "s-predicates have an *unaccusative* (in the sense of having an internal subject) Infl: the subject is base-generated internal to VP in [Spec, VP] and Infl does not assign a θ -role to [Spec, IP]" (p. 24).
 b. "i-predicates should be analyzed as analogues to control predicates. On this account, [...] i-predicates differ from s-predicates in that they have an Infl that assigns a θ -role to [Spec, IP]. This role has roughly the meaning 'has the property x ', where x is the property expressed by the predicate. The lexical NP in [Spec, IP] controls a PRO subject in [Spec, VP], which is assigned a θ -role by the verb" (pp. 25-26).

I am now ready to illustrate Diesing's account for the contrast in (287). By assumption (293), the BPS 'Italians' of the s-predicate 'available' can either remain at LF in its surface position [Spec, IP]

²² In subsection 3.1.1, I have suggested that Kratzer's strong assumption (236), that i-predicates lack a Davidsonian argument, could be replaced by the weaker assumption (239), that i-predicates have a Davidsonian argument too but that it is not temporally located. I think that the derivation of (292) from (236) just presented does not hold anymore once (236) is replaced by (239).

²³ Another possible argument in favor of Diesing's assumption (293) against Kratzer's assumption (292) has to do with the details of the compositional interpretation. Diesing assumes a semantics whereby the nuclear scope of Q-adverbs and the scope of the DEO are of type t . If these two scopes are both defined to be VP, then VP must be of type t . And in order for VPs to be of type t when their subject is generated outside of VP, we need a place holder in [Spec, VP].

or else be reconstructed into VP; hence, sentence (287a) with the s-predicate ‘available’ admits both LFs (294a) and (294a’).²⁴ Both LFs contain a matrix covert generic operator GEN, whose nuclear scope VP I have boldfaced. Again by assumption (293), the BPS ‘Italians’ of the i-predicate ‘italian’ cannot be reconstructed into VP, because it is base generated in [Spec, IP] from where it controls a null PRO in [Spec, VP]; hence, sentence (287b) with the i-predicate ‘tall’ only admits the LF (294b).



By assumption (291), the three LFs (294) yield the corresponding truth conditions (295). The BPS ‘firemen’ of the analogous LFs (294a) and (294b) sits in the restrictive clause of the generic operator and thus gets bound by it, yielding the truth conditions (295a) and (295b). The BPS ‘firemen’ of the LF (294b’) sits in the nuclear scope of the generic operator and thus gets bound by the DEO, yielding the truth conditions (295a’).

- (295) a. $GEN_{x,t}[[\text{fireman}]](x)[[\text{available}]](x,t)$.
 a' $GEN_t[\exists x[[\text{fireman}]](x) \wedge [[\text{available}]](x,t)]$.
 b. $GEN_x[[\text{fireman}]](x)[[\text{tall}]](x)$.

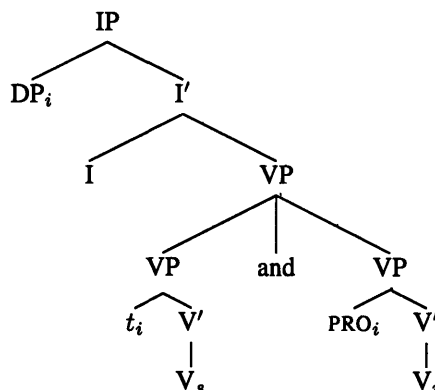
The truth-conditions (295b) correspond to the generic reading of the BPS of sentence (287b); since these are the only truth-conditions that we can derive for sentence (287b), then Diesing correctly predicts the BPS of this sentence to only admit the generic reading. The truth-conditions (295a) and (295a’) correspond respectively to the generic and the existential readings of the BPS of sentence (287a); hence, Diesing correctly predicts that the BPS of sentence (287a) admits both readings. The contrast in (287) is thus derived.

First argument against Diesing’s version of the syntactic approach By assumption (293), s- and i-predicates are selected by a different inflectional head. For the sake of explicitness, let me denote by Infl_s the “raising” inflectional head that selects s-predicates and by Infl_i the “control” inflectional head that selects i-predicates. Various authors have noted that this assumption (293) raises a problem for VP-coordination; see for instance Burton and Grimshaw (1992, p. 311). The core idea of the argument is as follows: assumption (293) combined with a plausible formulation of the coordinate structure constraint (henceforth: CSC) predicts the ungrammaticality of a coordinate

²⁴I assume that when the subject is reconstructed into [Spec, VP], it does not leave any trace in [Spec, IP], at least as far as semantics is concerned.

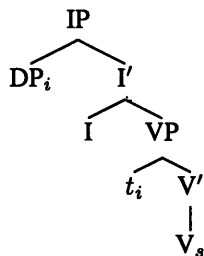
structure like (296), where a VP projected by an s-predicate V_s is conjoined with a VP projected by an i-predicate V_i ; yet, this coordinate structures are perfectly grammatical.

(296)

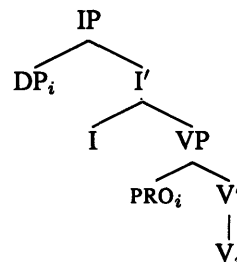


Here is a way to spell out the details of the argument. Take for instance the formulation of the CSC in Fox (2000, p. 50). It says that a necessary condition in order for the coordinate structure (296) to be grammatical is that its two components (297a) and (297b) both satisfy all grammatical constraints. Of course, the head Infl should be the same in both components (297a) and (297b), either $Infl_i$ or $Infl_s$. If it is $Infl_s$, then the component (297b) is ungrammatical, because $Infl_s$ does not assign any θ -role and therefore the DP_i in [Spec, IP] of (297b) receives no θ -role. If it is $Infl_i$, then the component (297a) is ungrammatical, because the chain $\langle DP_i, t_i \rangle$, receives two θ -roles, one assigned to t_i by V_s and a second one assigned to DP_i by $Infl_i$.

(297) a.



b.



To conclude the argument, we only need to exhibit examples of coordinate structures of the form (296) that sound fine. Some care is needed in constructing these examples, since we need to make sure that they involve VP-coordination rather than IP-coordination. One such example is (298), from Chierchia (1995, p. 221), building on van Valin (1986): the elided phrase is clearly the coordinated phrase in the antecedent sentence; since the tense is realized on 'did', it is implausible that the elided phrase could be a coordination of IPs; thus, plausibly the antecedent involves VP-coordination. The felicity of sentence (298) thus raises a problem for Diesing's assumption (293).

(298) Which of your colleagues came from America and moved to Russia?
Sue did \emptyset .

An alternative way of probing for VP-coordination comes from gapping. The subject quantifier 'every girl' of the first conjunct of sentence (299a) can bind the pronoun 'her' in the second conjunct, while this is not possible in sentence (299b). Furthermore, sentence (300a) allows the reading where negation has scope over disjunction, while sentence (300b) only admits the reading where negation has scope below negation. The facts about sentences (299b) and (300b) are straightforwardly accounted for by the assumption that these two sentences involve TP-coordination. In fact, the subject quantifier 'every girl' in [Spec, TP] of the first conjunct of sentence (299b) does not c -command the pronoun 'her' in [Spec, TP] of the second conjunct. Furthermore, negation heads a phrase NegP which immediately dominates VP so that TP-disjunction in sentence (300b) c -commands the negation of both conjuncts. Johnson (1996, 2000) and Lin (2000, 2001) suggest that the different behaviour of sentences (299a) and (300a) follows straightforwardly from the assumption that these sentences involve only one node T which selects as a complement a VP-coordination. In fact, assume that in

the case of sentence (299a) the verb ‘ate’ has moved ATB from both conjoined VPs to T and that the subject quantifier ‘every girl’ has moved from its base position inside the first conjoined VP to [Spec, TP], in order to satisfy EPP of T²⁵. This structure correctly predicts that the subject quantifier in [Spec, TP] *c*-commands the pronoun inside the second VP. Furthermore, assume that in the case of sentence (300a) negation is generated as the head of a phrase NegP which immediately *c*-commands the coordinated VPs. This structure correctly predicts that negation *c*-commands disjunction, yielding the attested reading of sentence (300a).

- (299) a. Not every girl_i ate a green banana, and her_i mother a ripe one.
 b. *Not every girl_i ate a green banana, and her_i mother ate a ripe one.
- (300) a. Kim didn’t play bingo or Sandy sit at home all night.
 = ¬[[play]](*kim*) ∧ ¬[[sit]](*sandy*)
 b. Kim didn’t play bingo or Sandy didn’t sit at home all night.
 = ¬[[play]](*k*) ∨ ¬[[sit]](*s*)

This analysis of the minimal pairs (299) and (300) implies that the subjects ‘her mother’ and ‘Sandy’ sit in [Spec, TP] in the case of sentences (299b) and (300b) but sit in [Spec, VP] in the case of sentences (299a) and (300a). Therefore, Diesing’s assumption (293) predicts that the *s*-predicates ‘eat’ and ‘sit’ could be replaced with *i*-predicates in sentences (299b) and (300b) but not in sentences (299a) and (300a). This prediction is not borne out, since sentences (301a) and (302a) are fine.

- (301) a. Not every girl_i loves an old man, and her_i mother a young man.
 b. *Not every girl_i loves an old man, and her_i mother loves a young man.
- (302) a. Kim doesn’t love Mary or Sandy love John.
 b. Kim doesn’t love Mary or Sandy doesn’t love John.

Second argument against Diesing’s version of the syntactic approach So far, I have tried to argue against Diesing’s crucial assumption (293) by looking at the case of coordination. As noted above, the intended effect of this assumption is that of blocking reconstruction of the subject of an *i*-predicate into VP. Thus, a somewhat more direct way to argue against Diesing’s proposal is to look for diagnostics for reconstruction and show that the diagnostics do not differentiate between subjects of *s*- and *i*-predicates. Here is a way to pursue this strategy. Various arguments have been provided to show that subjects can be lowered down to a position underneath negation; see for instance Johnson and Tamioka (1997) for one such argument. As stated in (303), these arguments do not distinguish between *s*- and *i*-predicates.

- (303) Subjects of both *i*- and *s*-predicates can be QLed at LF to a position lower than NegP.

This is not enough yet to threaten Diesing’s proposal: it could be the case that the position to which the subject may be QLed, although lower than NegP, is still high enough to be outside of the nuclear scope of the generic operator. To complete the argument, consider a sentence such as (304), with negation and a Q-adverb. This sentence does admit the reading that almost for every student it is the case that he does not come from Asia.

- (304) MIT students usually do not come from Asia.

In order to derive the intended reading of sentence (304), we need to assume that the nuclear scope of ‘usually’ contains NegP. Plausibly, this is a general property of Q-adverbs, as stated in (305). By (303), subjects of *i*-predicates may be lowered to a position inside NegP; by (305), NegP can sit within the nuclear scope of Q-adverbs; in conclusion, subjects of *i*-predicates may be lowered inside the nuclear scope of Q-adverb, contrary to what predicted by Diesing’s assumption (293).

- (305) NegP may sit inside the nuclear scope of a Q-adverb.

²⁵See Lin (2001) for an argument that this extraction of the subject from the first conjunct of a coordinate structure is compatible with Fox’s (2000; p. 50) formulation of the CSC.

In conclusion, it does not seem to me possible to account for the lack of the existential reading of BPSs of *i*-predicates by assuming that these subjects must always sit outside of the nuclear scope of the generic operator at LF. In other words, constraints on subject reconstruction are orthogonal to the distribution of the existential reading of BPSs.

The information structure approach Jäger (2001) and Cohen and Erteschik-Shir (2002) develop an alternative account for the distribution of the existential reading of BPSs, that makes use of constraints on information structure rather than on syntactic assumptions on reconstruction. The two accounts seem to me almost identical; thus I will not distinguish between the two, and rather collectively refer to them as the JCS's account. The crucial notion of information structure used in the JCS account is that of TOPIC. The main assumptions on topics needed by this account are summarized in (306).²⁶

- (306)
- a. Every atomic clause must have a topic.
 - b. In the case of *s*-predicates, the Davidsonian argument can serve as a topic; this option is not available for the case of *i*-predicates.
 - c. Q-adverbs unselectively bind all topics in their scope.

The availability of the existential reading of the BPS 'firemen' of the *s*-predicate 'available' of sentence (287a) can be accounted for as follows: since 'available' is an *s*-predicate, its Davidsonian argument can serve as the topic of the sentence, by assumption (306b); thus, the constraint (306a) that each sentence has a topic is satisfied without any need for the BPS to be marked as topic; since the BPS is not topical, it can get a non-specific existential reading. The lack of the existential reading of the BPS 'firemen' of the *i*-predicate 'tall' of sentence (287b) can be accounted for as follows: since 'tall' is an *i*-predicate, there is no Davidsonian argument around that can serve as the topic of the sentence, by assumption (306b); thus, the constraint in (306a) that each sentence has a topic must be satisfied by marking the BPS as topic; being topical, the BPS can only get the generic reading, by assumption (306c). The JCS's account just sketched seems to me to face major empirical problems. In the rest of this subsection, I discuss these problems by concentrating on Jäger (2001); the same issues seem to me to hold against Cohen and Erteschik-Shir (2002).

²⁶ Jäger (2001) and Cohen and Erteschik-Shir (2002) spell out the crucial assumption (306b) in two different ways. Cohen and Erteschik-Shir adopt Kratzer's assumption (236) that *i*-predicates lack a Davidsonian argument, from which (306b) follows straightforwardly. Jäger's take on the issue is more elaborate. His starting point is the observation that also BPSs of *s*-level statives often lack the existential reading, an issue I will come back to in subsection 3.6.4. In order to capture this fact, he formulates assumption (306b) without distinguishing between *i*- and *s*-level statives with respect to the topicality of their Davidsonian argument, as in (i).

- (i) "Events [namely the Davidsonian arguments of eventives] but not states [namely the Davidsonian arguments of statives] can function as a topic" (p. 122).

But why should (i) hold? Jäger assumes that "all predicates [both eventives and statives] have a Davidsonian argument" (p. 102) and that in principle all Davidsonian arguments (both states and events) could be topics. He then derives his assumption (i) by means of the piece of reasoning in (ii).

- (ii)
 - a. Davidsonian arguments can be topic only if they can be bridged: "Davidsonian arguments are always novel, thus to be topics, they have to be linked to the preceding discourse via bridging" (p. 122).
 - b. Davidsonian arguments of (both *i*- and *s*-level) statives are "of a different sort than events proper", namely "are not localized in space" (p. 122), while events are.
 - c. "events [being localized in space] easily undergo bridging via the relation of local nearness, while states [being not spatially localized] cannot be bridged in this way" (p. 122).
 - d. In conclusion, since only events can be bridged while states cannot be bridged, then "in effect events but not states can function as a topic", as stated in (i).

Yet, the reasoning in (ii) crucially relies on the implicit assumption that bridging can only hold "via the relation of local nearness". Why is this so? Why can't bridging also happen, say, via the relation of temporal nearness? If the latter were indeed the case, then Davidsonian arguments of non-permanent statives, being temporally located, should be able to undergo bridging and thus to function as topic. Thus, it seems to me that Jäger's account for the parallelism between *i*-level and *s*-level statives comes at the price of the bare stipulation that bridging can only happen through spatial nearness and not through temporal nearness.

First argument against the information structure approach So far, I have only considered sentences with an intransitive predicate, and thus a single argument, namely the subject. As Jäger himself notes, his proposal predicts that things should be very different as soon as we add another argument besides the subject:

- (307) “Non-subject arguments (even implicit ones) can serve as topic and thus save a weak reading of the subject of a stative predicate” (p. 122).

Jäger suggests that this prediction is borne out for instance by the contrast in (308): the BPS of sentence (308a) lacks the existential reading while the BPS of sentence (308b) readily admit it. He suggests that we can make sense of this fact in terms of (307): assume that the object ‘that house’, being specific, can be topic; thus, the subject in (308b) is relieved from the duty of serving as topic; hence, this subject can be construed existentially. With Jäger’s own words: “we predict that the possibility of a weak construal of some subjects does not depend solely on the main predicate but on the other arguments as well. If the object is a topic, the discourse linking principle [namely (306a)] is fulfilled and the sentence [with an existential construal of the subject] is felicitous” (p. 120).

- (308) a. Tycoons own banks. * \exists -BPS
 b. Tycoons own that house. $\sqrt{\exists}$ -BPS

Yet, Jäger’s prediction (307) seems to me quite wrong, as can be seen in a number of ways. Consider the minimal pair in (309): since sentence (309b) contains the specific object ‘Pavarotti’ which can act as a topic, Jäger predicts that the BPS ‘Italians’ in (309b) should admit the existential reading, contrary to the BPS in (309a). This prediction seems to me wrong: there is no difference between the BPS of (309a) and that of (309b). The same point can be trivially repeated with all classical examples of transitive i-predicates.

- (309) a. Italians love good singers.
 b. Italians love Pavarotti.

Consider the minimal pair in (310): since sentence (310b) contains the argument ‘Michael Jordan’ which can act as a topic, Jäger predicts that the BPS ‘firemen’ in (310b) should admit the existential reading, contrary to the BPS in (310a). This prediction seems to me wrong: there is no difference between the BPS of (310a) and that of (310b). The same point can be made with all adjectival i-predicates.

- (310) a. Firemen are tall.
 b. Firemen are taller than Michael Jordan.

My conclusion is that Jäger’s prediction (307) is not quite right: by and large, the availability of the existential reading of the subject does not in any way depend on whether there are other arguments around that can serve as topic. And the few cases where the prediction (307) seems to hold, such as (308), are better analyzed as special cases, for which something *ad hoc* needs to be said.²⁷

Second argument against the information structure approach²⁸ Diesing and Chierchia account for the lack of the existential reading of BPSs of i-predicates by means of assumptions on subject reconstruction. I have argued against their approach by suggesting that it is incompatible with other

²⁷Jäger gives two more examples that he considers parallel to (308), which I repeat below, together with Jäger’s judgments.

- (i) a. Monkeys live in trees. (Generic only)
 b. Monkeys live in that tree. (Existential possible)
 (ii) a. Presidents are similar to senators. (Generic only)
 b. Presidents are similar to these senators. (Existential possible)

The example in (i) seem to me not bear on the issue, since ‘live’ is not an i-predicate nor a stative. The example in (ii) is dubious at best: my informants consider the existential reading of the BPS of sentence (iib) not available (Ezra Keshet and Jennifer Michaels (p.c.)). This latter point shows that the availability of existential BPSs of transitive i-predicates with specific objects is really very tightly constrained, contrary to the prediction (307).

²⁸The material in this paragraph was prompted by a discussion with Angelika Kratzer.

existing applications of the theory of reconstruction. Jäger and Cohen and Erteschik-Shir account for the lack of the existential reading of BPSs of *i*-predicates by means of assumptions on information structure. Let me now argue against their approach by suggesting that it is incompatible with other existing applications of the theory of information structure, such as that described in Heycock (1994). Heycock provides an account for a very interesting contrast between the interpretation of Japanese ‘ga’-marked subjects in matrix and embedded clauses in terms of information structure. The matrix sentence (311a), with the *s*-predicate ‘kita’ (‘come’), admits two readings: “on the NEUTRAL DESCRIPTION READING [henceforth: NDR], this sentence is a straightforward announcement of the event of John’s arrival; on the EXHAUSTIVE LISTING READING [henceforth: ELR], it means something like ‘of all the people salient at this point in the discourse, it was John who came’”. The matrix sentence (311b), with the *i*-predicate ‘kasikoi’ (‘smart’), is not ambiguous: “if the predicate is *i*-level [...] a subject marked with ‘ga’ can only receive the ELR. [(311b)] is such an example: it is interpreted roughly as ‘of all the people salient at this point in the discourse, it is John who is smart’. The NDR is not available for this sentence.” This restriction on the NDR of ‘ga’-marked subjects of *i*-predicates disappears in embedded clauses: no matter whether the embedded clause contains an *i*- or an *s*-predicate, both the NDR and the ELR are available for the ‘ga’-marked subject of the embedded clause, as in the case of (312).

(311) Matrix sentences:

- | | |
|---|--------------------|
| a. John-ga kita.
John-GA came.
‘John/JOHN came’ | [both NDR and ELR] |
| b. John-ga kasikoi.
John-GA smart
‘JOHN is smart’ | [only ELR, no NDR] |

(312) Embedded sentences:

- | | |
|--|--------------------|
| a. Mary-wa [John-ga kita koto-o] wasurete-ita.
Mary-WA [John-GA came fact-ACC] forgot.
‘Mary forgot that John/JOHN came’ | [both NDR and ELR] |
| b. Mary-wa [John-ga kasikoi koto-o] wasurete-ita.
Mary-WA [John-GA smart fact-ACC] forgot.
‘Mary forgot that John/JOHN is smart’ | [both NDR and ELR] |

In conclusion, the interpretations available for ‘ga’-marked subjects of matrix *i*-predicates are restricted compared to matrix *s*-predicates: only the ELR but not the NDR is available; the restriction disappears in embedded contexts. Heycock’s account for the facts in (311)-(312) is cast within the theory of information structure. More precisely, Heycock’s core assumptions are (313a)-(313c).

- (313) a. i. A *matrix* sentence must have a topic²⁹ and a focus.
ii. An *embedded* sentence need not have both a topic and a focus.
- b. “In sentences with *s*-predicates, the Davidsonian event argument referring to the slice of time and space at which the event takes place is always available as a topic. [...] If the predicate is an *i*-predicate, however, there is no such event argument”.
- c. i. ‘ga’ is a nominative marker, with no information structure import;
ii. an argument is a topic iff it is marked with ‘wa’;
iii. a predicate can be topic without being marked with ‘wa’.

Here is how Heycock uses these assumptions (313a)-(313c) to account for the facts in (311)-(312). Consider first the case of the matrix sentence (311a) with the *s*-predicate ‘kita’ (‘come’). By assumption (313b), the Davidsonian argument of the *s*-predicate can count as a topic. Thus, the entire

²⁹Heycock uses the term “link” instead of “topic”, but makes it explicit that “the link corresponds to a large extent to what in some frameworks is called the topic”; I stick to the term ‘topic’ here, and always replace ‘link’ with ‘topic’ in all my quotes from Heycock.

sentence can be marked as focus, as in the Information Structure in (314a) which corresponds to the NDR. Suppose instead we decide not to mark as topic the Davidsonian argument of the s-predicate. By assumption (313ai), we have to mark as topic some other constituent. By assumption (313cii), the subject cannot be marked as topic, because it has been marked with 'ga' rather than 'wa'. By assumption (313ciii), the predicate can be marked as topic, and is thus the only option. By assumption (313ai), we also have to mark as focus some constituent in the sentence and the only option is to mark as focus the subject. Thus, we derive the Information Structure in (314b), which corresponds to the ELR. Consider next the case of the matrix sentence (311b) with the i-predicate 'kasikoi' ('smart'). By assumption (313b) that i-predicates lack a Davidsonian argument, the requirement (313ai) that matrix sentences have a topic cannot be satisfied by the Information Structure in (314a) in the case of i-predicates. Hence, the only Information Structure available for the matrix sentence (311b) with the i-predicate 'kasikoi' ('smart') is the one in (314b), thus accounting for the fact that this sentence only admits the ELR. Of course, this contrast between i- and s-predicates disappears in embedded clauses: by assumption (313aii), there is no requirement for the embedded clauses in (312) to contain a topic, and thus the option in (314a) of focus marking the entire embedded clause is available both in the case of s- and i-predicates.

- (314) a. [_F Subj VP]. ⇒ NDR
 with the Davidsonian argument marked as topic.
 b. [[_F Subj] [_T VP]]. ⇒ ELR
 with the Davidsonian argument unmarked.

Heycock's account for restrictions on the NDR/ELR ambiguity for 'ga'-marked subjects of matrix predicates looks attractively similar to JCS's account for restrictions on the existential/generic ambiguity for BPSs reviewed above. As a matter of fact, Heycock's assumptions (313a) and (313b) correspond to assumptions (306a) and (306b) respectively, and the two parallel pairs of assumptions do indeed play a very similar role in the two accounts. Thus, these two accounts seem to outline a unified approach to i-predicates based on a few core assumptions on information structure. Yet, I want to argue that the convergence between the two proposals is only apparent. More precisely, I will argue for claim (315). To the extent that Heycock's proposal that the NDR/ELR ambiguity is governed by information structure constraints is correct, it follows that JCS are wrong in claiming that the existential/generic ambiguity is governed by those same constraints too.

- (315) The NDR/ELR ambiguity for 'ga'-marked subjects and the existential/generic ambiguity for BPSs have different empirical properties, and thus cannot be subsumed under a unified theoretical framework.

Heycock notes that the NDR/ELR ambiguity is affected by the difference between matrix and embedded clauses. But the existential/generic ambiguity for BPSs is not affected by that difference: the BPS 'firemen' cannot be construed as existential neither in the matrix clauses (316a) nor in the embedded clauses (316b). This difference with respect to the matrix/embedded distinction provides an argument in favor of claim (315) that the restrictions on the NDR/ELR ambiguity and the restrictions on the existential/generic ambiguity have different empirical properties. This means that it is not possible to account for both phenomena in terms of constraints on information structure. In fact, assume that Heycock is on the right track in claiming that the NDR/ELR should be accounted for in terms of information structure. This means that embedded clauses must be relieved from the requirement of having a topic, as stated in (313aii). But once we import this assumption (313aii) within JCS's account, we make the incorrect prediction that BPSs of i-predicates should admit the existential reading in embedded clauses: if embedded clauses can do without a topic, nothing forces their BPSs to be topics and thus nothing prevents them from having an existential reading.

- (316) a. Firemen are tall / intelligent / altruistic / Italian...
 b. John thinks that firemen are tall / intelligent / altruistic / Italian...

Heycock notes that "there is the possibility that some topic can be recovered from the context, even in the case of an i-predicate. In this case we predict that the interpretation need not be one of narrow focus on the subject [as in the Information Structure (314b), which corresponds to the

ELR], since the entire sentence could constitute the focus [as in the Information Structure (314a), which corresponds to the NDR]. This prediction is borne out. Consider for example the exchange in [(317)]. The sentences in [(317B)] do not have to be interpreted with narrow focus on the subject [as in the ELR]. Rather the question provides ‘problems with B’s new job’ as the topic, leaving open the possibility that all the material in the response will be new information to be filed under this address.”

- (317) A: Atarasii sigoto-no mondai-wa nan desu ka?
 new work-GEN problem-WA what be QU
 ‘What’s the problem with your new job?’
 B: Ofisu-ga tiisaisi, kyuuryoo-ga yasuisi, uwayaku-ga hidoi desu.
 office-GA small-and, pay-GA low-and, boss-GA terrible be
 ‘The office is small, the pay is low, and the boss is terrible’

But the existential/generic ambiguity for BPSs is not affected by the presence of other topics provided by the discourse: as shown in (318), the BPSs ‘Italians’ of the *i*-predicate ‘lazy’ of sentence (318B) does not in any way admit the existential reading, despite the fact that, by parity of reasoning with Heycock’s example (317), one might conceive the preceding question (318A) as providing ‘problems with Italy’ as the topic, thus reliving the BPS from the duty of serving as topic. This difference with respect to sensitivity to discourse topics provides another argument in favor of claim (315) that the restrictions on the NDR/ELR ambiguity and the restrictions on the existential/generic ambiguity of BPSs have different empirical properties. Once more, this means that it is not possible to account for both phenomena in terms of constraints on information structure. In fact, assume that Heycock is on the right track in claiming that a topic can be provided by discourse. Once we import this assumption within JCS’s account, we make the incorrect prediction that existential BPSs of *i*-predicates should be available as soon as the context provides an implicit topic.

- (318) A: What’s the problem with Italy?
 B: The economy is going down and Italians are lazy.

The example in (319) sharpens this point. Heycock notes that “the exchange in [(319)] is felicitous because the question supplies ‘the best thing about New Haven’ as the topic. The predicate in [(319B)] is clearly individual-level; furthermore, there is no any reason to suppose that any ‘coercion’ into a stage-level reading is taking place. In [(319B)] the ‘ga’-marked subjects are clearly part of the focus”. Thus, JCS’s account would predict the BPS ‘buildings’ of the corresponding English paraphrase to allow the existential reading, which is wrong.

- (319) A: Sorejaa, New Haven-de ichiban ii mono-wa nani?
 so, New Haven-in most good thing-WA what
 ‘So what’s the best thing about New Haven?’
 B: Tatemono -ga kirei da.
 building-GA beautiful be
 ‘Buildings are beautiful’

A digression on Heycock’s pattern (311)-(312) Before I leave this discussion, I would like to quickly address Haycock’s pattern (311)-(312) from the perspective of the theory of *i*-predicates that I am defending in this work. I am not in a position to offer a detailed account, but would like at least to point out a possible way to go. I have repeated in (320a) and (320b) the relevant data concerning matrix sentences: matrix ‘ga’-marked subjects of *s*-predicates admit both the NDR and the ELR; matrix ‘ga’-marked subjects of *i*-predicates only admit the ELR. Let me consider next the case of matrix ‘ga’-marked subjects of habitual *s*-predicates: my Japanese informants (Junri Shimada and Yasutada Sudo (p.c.)) report that the ‘ga’-marked subject of the matrix habitual *s*-predicate in (320c) behaves exactly as the ‘ga’-marked subject of the matrix *i*-predicate in (320b), namely it only allows the ELR.

- (320) a. John-ga kita.
John-GA came.
'John/JOHN came' [both NDR and ELR]
- b. John-ga kasikoi.
John-GA smart
'JOHN is smart' [only ELR, no NDR]
- c. John-ga mainichi sake-o nomu.
John-GA every day sake-ACC drink-NON-PAST
'JOHN usually drinks sake' [only ELR, no NDR]

The core tenet of my proposal is that *i*-predicates have Davidsonian arguments and that their Davidsonian arguments do not have any special ontological properties. What's special about a sentence containing an *i*-predicate is that, in order for the sentence not to trigger a mismatching implicature, the Davidsonian argument of the *i*-predicate must be bound by the generic operator. Thus, the proper truth conditions of sentence (321a) = (320b) are those in (321b), as in Chierchia (1995). Note that the truth conditions in (321b) are formally analogous to the standard truth conditions (322b) for sentence (322a)=(320c) with the habitual *s*-predicate 'drinks sake'.

- (321) a. John is smart.
b. $GEN_t[C(t)][[smart]](j, t)$.
- (322) a. John drinks sake.
b. $GEN_t[C(t)][[drink-sake]](j, t)$.

Given the strict analogy between (321) and (322), whatever accounts for the lack of the NDR in the case of the 'ga'-marked subject of the habitual *s*-predicate in (320c) will plausibly account also for the lack of the NDR in the case of the 'ga'-marked subject of the *i*-predicate in (320b).

3.3.2 An account based on blind and mandatory mismatching implicatures

In this subsection, I introduce some standard assumptions on existential BPSs, independently of *i*-predicates. On the background of these assumptions, I then present my account for the contrast in (287). Finally, I discuss its extension to plain indefinites and 'there'-construction.

Preliminaries Let me start by presenting my assumptions on existential BPs. Carlson (1977) points out that existential BPs always take the narrowest possible scope with respect to other scope bearing elements in the sentence. Some of the examples he uses to make this point are quoted in (323)-(325). He comments on these examples as follows: "In the (a) versions, with the singular, the existential quantifier is interpreted as being outside the scope of the time adverbial, and thus it is the same object that is spoken of at all relevant points of time. In the (b) versions, however, the existential quantifier is interpreted as being within the scope of the adverbial, so the objects referred to may be different from time to time."

- (323) a. A dog hung around my house all last week.
b. Dogs hung around my house all last week.
- (324) a. A cat has been here since the Vikings landed.
b. Cats have been here since the Vikings landed.
- (325) a. A tyrant ruled Wallachia for 250 years.
b. Tyrants ruled Wallachia for 250 years.

One way of accounting for the generalization that existential BPs have narrow scope is through assumptions (326), from Diesing (1992) building on Heim (1988) and Kamp (1981). By (326a), the BPS 'dogs' of sentence (323b) introduces a free variable that ranges over dogs. This variable needs to be bound at LF and the DEO in (326b) is the only available binder. Since by (326c) the DEO has the smallest possible scope, the time adverbial 'all last week' must scope above it. In conclusion, the existential BPS 'dogs' can only have narrow scope w.r.t. the time adverbial 'all last week'.

- (326) a. A BP $[_{DP} \emptyset N]$ has no quantificational force on its own: it simply introduces a free variable x that ranges over $[N]$.
- b. There is a default existential operator (henceforth: DEO) that binds every free variable in its scope.
- c. The DEO has the narrowest possible scope: let's say it has VP scope, where VP is the smallest scope site (the smallest constituent of type t).

Another way of accounting for the generalization that existential BPs have narrow scope is thorough assumptions (327), from Chierchia (1998). By assumption (327a), a type (or sort) mismatch arises when the BPS 'dogs' is applied to the predicate 'hung around' in sentence (323b), since the former denotes a kind while the latter denotes a property of regular individuals. By assumption (327b), this type mismatch is resolved through the type shifter DKP, that raises the type (or sort) of the predicate from regular individuals to kinds. By assumption (327c), DKP is a last resource option, and is thus triggered only in the most embedded position. In conclusion, the existential BPS 'dogs' can only have narrow scope w.r.t. the time adverbial 'all last week'.

- (327) a. Bare plurals denote kinds, after Carlson (1977).
- b. The DERIVED KIND PREDICATION operator $\langle \cdot \rangle_{DKP}$ shifts a predicate P of common individuals into a predicate of kinds in the following way:

$$\langle P \rangle_{DKP} \doteq \lambda x_k . \exists y [y \leq x_k \wedge P(y)]$$

where x_k is a variable over kinds and the relation $d \leq k$ between a common individual d and a kind k holds iff d is an instance of the kind k .

- c. DKP is a type shifter and thus is only applied as a last resource.

I can develop my account on the background of either of these frameworks; for concreteness, I'll adopt the Heim-Diesing's framework (326). With this bit of preliminaries in place, let's now turn to the case of sentence (287a), with the BPS 'firemen' and the s-predicate 'available'. According to the Heim-Diesing framework (326), the existential reading of the BPS of this sentence corresponds to the LF in (328). The generic or existential operator binds the time argument of the predicate, yielding the habitual or the episodic reading, respectively. A contextually supplied predicate over time C restricts the tense operator GEN_t or \exists_t . The BPS has been reconstructed into its base position within VP, whereby the free variable x it introduces can get bound by the DEO, yielding the existential reading of the BPS.

$$(328) \quad [[\left\{ \begin{array}{c} GEN_t \\ \exists_t \end{array} \right\} C(t)]]_{VP} \text{ firemen}(x) \text{ available}(t)]]$$

The truth conditions which correspond to the LF (328) are those in (329). These truth-conditions say that "for some/every time t which satisfies the restrictor \tilde{C} , there is a fireman x who is available at t ." Note that the tense operator GEN_t/\exists_t in (329) has wide scope over the existential quantifier over firemen \exists_x . That this is the only possible scope configuration follows straightforwardly: since the existential BP only arises through the DEO, since the DEO has VP-scope and since VP is the smallest scope site, then GEN_t/\exists_t must scope above \exists_x .

$$(329) \quad \lambda w . \left\{ \begin{array}{c} GEN_t \\ \exists_t \end{array} \right\} [\tilde{C}^w(t)] [\exists_x [[\text{firemen}]^w(x) \wedge [\text{available}]^w(x, t)]]]$$

As a final point, note that the LF (328) contains the restrictor C while the truth conditions (329) contain a different restrictor \tilde{C} . Here is why. By assumption (247), the formula $[\text{available}]^w(x, t)$ presupposes that x is alive at t . Let π be the presupposition that we get by letting this presupposition pass through the existential quantifier \exists_x . By assumption (248), this presupposition π must be added via conjunction to the restrictive clause C of the tense operator. Thus, I am assuming that \tilde{C} is the conjunction of C with this presupposition π . Since I am considering an arbitrary C to start with, it does not really matter what this presupposition π looks like.

The account In order for a BPS to get an existential reading, the variable it introduces must be bound by the DEO; in order for this binding to hold, the variable must fall within the DEO's scope, namely VP; in order for this configuration to hold, the BPS must be reconstructed within VP. Diesing straightforwardly accounts for the lack of the existential reading of BPSs of i-predicates by assuming that subjects of i-predicates cannot be reconstructed into VP. This option is of course not available to me, given my tenet that there are no grammatical differences between s- and i-predicates. Thus, let me assume that the same LFs (328) and the same truth conditions (329) just derived for sentence (287a) with the s-predicate 'available' are available also for sentence (287b) with the i-predicate 'tall'. I have written them down in (330) and (331), for the two cases where the time argument of the i-predicate is bound by the generic or by the existential operator. Thus, the goal of accounting for the lack of the existential reading of the BPS 'firemen' of the i-predicate 'tall' of sentence (287b) boils down to the goal of ruling out both truth conditions φ and φ' in (330b) and (331b).

$$(330) \quad \text{a. } [[\text{GEN}_t \text{ C}(t)]]_{[\text{VP} \text{ firemen}(x) \text{ tall}(t)]}] \\ \text{b. } \varphi = \lambda w . \text{GEN}_t [\tilde{\text{C}}(t)] [\exists_x [[\text{firemen}]^w(x) \wedge [[\text{tall}]^w(x, t)]]].$$

$$(331) \quad \text{a. } [[\exists_t \text{ C}(t)]]_{[\text{VP} \text{ firemen}(x) \text{ tall}(t)]}] \\ \text{b. } \varphi' = \lambda w . \exists_t [\tilde{\text{C}}^w(t)] [\exists_x [[\text{firemen}]^w(x) \wedge [[\text{tall}]^w(x, t)]]].$$

Let me start with the case of φ in (330b). Consider the LF in (332a), which is identical to the one in (330a) but with the BPS 'firemen' replaced by the definite description 'the fireman such and such' (here abbreviated as 'the fireman P '). On the background of the two assumptions (247) and (248), the LF (332a) yields the truth conditions ψ_P in (332b), where I have used the shorthand d_P^w for the individual $[[\text{the fireman } P]^w]$. These truth conditions say that "at each time t in C^w at which the fireman such that P is alive, he is tall at t ."

$$(332) \quad \text{a. } [[\text{GEN}_t \text{ C}(t)]]_{[\text{VP} \text{ the fireman } P \text{ tall}(t)]}] \\ \text{b. } \psi_P \doteq \lambda w . \text{GEN}_t [\text{C}^w(t) \wedge \text{in}^w(d_P^w, t)] [[[\text{tall}]^w(d_P^w, t)]].$$

As noted above, the LF (332a) is obtained from the LF (330a) by replacing 'firemen' with 'the fireman such and such'. Let me consider the slight variant of assumption (110) stated in (333), namely that the definite description 'the fireman such and such' is a Horn-mate of both the indefinite 'a/some fireman' and the bare plural 'firemen'. Since the LF (332a) is obtained from the LF (330a) by replacing the scalar item 'firemen' with its Horn-mate 'the fireman P ', then ψ_P is a scalar alternative of φ , according to the definition of scalar alternatives (74).³⁰

(333) 'firemen', 'the fireman such and such' is a Horn-scale.

Let ψ be as in (334), which says that "there is a fireman x who is always tall throughout the portion of his lifespan which is in C ". Thus, ψ is as φ in (330b) but with the opposite scope $\exists_x > \text{GEN}_t$.

$$(334) \quad \psi = \lambda w . \exists_x [[\text{fireman}]^w(x) \wedge \text{GEN}_t [\text{C}^w(t) \wedge \text{in}^w(x, t)] [[[\text{tall}]^w(x, t)]]].$$

³⁰Furthermore, since ψ_P in (332b) is not weaker than φ in (330b), then ψ_P is a scalar alternative of φ also according to Fox's alternative definition (124) of the set of scalar alternatives. The two assumptions (247) and (248) clearly play a crucial role in deriving the truth conditions ψ_P in (332); let me discuss an alternative way to derive these same truth conditions without using the two assumptions (247) and (248) and explain why I don't go that way. Consider the following LF (i). This LF differs from the LF (330a) of our sentence in two respects: first, the BPS 'firemen' has been replaced by the definite 'the fireman such and such'; second, the restrictive clause $\text{C}(t)$ has been replaced by a different one, namely $\text{C}(t) \wedge \text{in}(d_P, t)$. Of course, this LF (i) yields the truth conditions ψ_P in (332b) straightforwardly, without any need for the two assumptions (247) and (248).

(i) $[[\text{GEN}_t [\text{C}(t) \wedge \text{in}(d_P, t)]]] [\text{the fireman } P \text{ tall}(t)]]$

The reason why I do not follow this way, is that the LF (i) is *not* a scalar alternative of the LF (330a) of our sentence, according to the alternative definition (124) of scalar alternatives that I am adopting here, from Fox (2007a, ft. 35). In fact, no matter whether the replacement of the restrictive clause $\text{C}(t)$ with $\text{C}(t) \wedge \text{in}(d_P, t)$ is performed before or after the replacement of the BPS with the definite, it leads to a weaker formula and thus is not licit according to definition (124). Thus, I need the restriction $\text{in}(d_P, t)$ to slip into the restrictive clause in some other way. The two assumptions (247) and (248) do the trick. This trick was suggested to me by Irene Heim (p.c.).

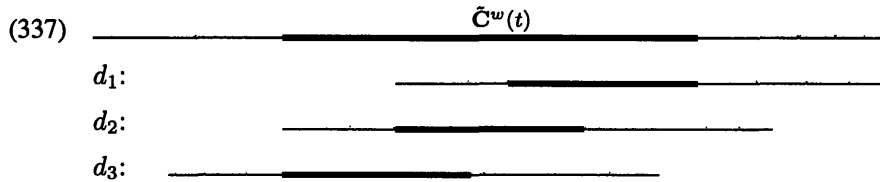
The logical equivalence in (335) holds:³¹ the negation of all the alternatives ψ_P 's in (332) is equivalent to the negation of ψ in (334), which says that "there is no fireman who is tall throughout the entire portion of his life span which is in C ."

$$(335) \quad \bigwedge_P \neg\psi_P = \neg\psi$$

Let me now argue that the blind strengthened meaning of φ in (330) is the one in (336), namely that all the ψ_P 's in (332) are excludable w.r.t. φ . Given the logical equivalence in (335), it suffices to exhibit a world where $\varphi \wedge \neg\psi$ is true.

$$(336) \quad \text{EXH}(\varphi) = \varphi \wedge \underbrace{\bigwedge_P \neg\psi_P}_{\neg\psi}$$

Consider the possible world w represented in (337). Suppose that there are three firemen in w , namely d_1 , d_2 and d_3 . Their life span is represented with a thin segment. The thickened portion of each segment represents the portion of each life span throughout which the individual is tall in w . The long thin line at the top of the picture is the time axis T ; its thickened portion is the restrictive clause of formula φ , i.e. $\tilde{C}^w(t)$. Note that φ is true in this world w , because for each time t in the set \tilde{C}^w , either d_1 or d_2 or d_3 is tall at t . Note furthermore that ψ is false in w , since no one of the three firemen d_1 , d_2 , d_3 is tall throughout the entire portion of his life span which is in \tilde{C}^w . Thus, $\text{EXH}(\varphi) = \varphi \wedge \neg\psi$ is not a logical contradiction.



As stated in (338), the blind strengthened meaning $\text{EXH}(\varphi) = \varphi \wedge \neg\psi$ just computed is a contradiction given the common knowledge \mathcal{W}_{ck} in (227) that 'tall' is a permanent property, precisely because no world such as the one in (337), in which firemen are tall only throughout a portion of their life span, can ever belong to \mathcal{W}_{ck} . In fact, suppose that φ is true in a world $w \in \mathcal{W}_{\text{ck}}$. Hence, there must exist at least one fireman in w who is tall at at least one instant in w , say John. Since w is compatible with the common knowledge \mathcal{W}_{ck} in (227), John is tall throughout his entire life span in w and ψ has got to be true too in w .

$$(338) \quad \text{EXH}(\varphi) \wedge \mathcal{W}_{\text{ck}} = \varphi \wedge \neg\psi \wedge \mathcal{W}_{\text{ck}} = \emptyset$$

Let me conclude. The existential reading of the BPS 'firemen' of the i-predicate 'tall' of sentence (287b) could in principle correspond to one of the two truth conditions φ or φ' in (330b) and (331b), depending on whether the time argument of the i-predicate is bound by the generic or by the existential operator. The first option φ in (330b) leads to the blind strengthened meaning (336), which contradicts the common knowledge in (227) and is thus ruled out by the MH (227). The second option φ' in (331b) can easily be shown to lead to a strengthened meaning which contradicts common knowledge too and is thus likewise ruled out by the MH.³² In conclusion, the lack of the existential reading of the BPS 'firemen' of the i-predicate 'tall' of sentence (287b) follows from the fact that every LF which would correspond to such a reading is ruled out by the MH.

³¹The reasoning is identical to that in footnote 16.

³²Here is a sketch of the reasoning. From the assumption (254) that \exists_t and \forall_t are Horn-mates together with the assumption (333) that 'firemen' and 'the fireman such and such' are Horn-mates too, it follows that the set of scalar alternatives $\text{Alt}(\varphi')$ of φ' in (331b) contains all the alternatives listed in (i):

- (i) a. the alternative ϕ obtained by replacing \exists_t with \forall_t ;
- b. all the alternatives ϕ_P 's obtained by replacing the BPS 'firemen' with the definite 'the fireman such and such';
- c. all the alternatives ψ_P 's in (332) obtained by replacing both \exists_t with \forall_t and the BPS with the definite.

It is easy to check that the set of excludable alternatives w.r.t. φ' according to Fox's (2007) definition (106) is $\text{Excl}(\varphi') = \{\phi, \psi_P \mid P\}$. Hence truth conditions φ' in (331b) are ruled out in exactly the same way I have ruled out above truth conditions φ in (330b).

First remark: indefinites Let me consider again sentence (287b) repeated below in (339a), together with the variant (339b), with the indefinite ‘some fireman’ instead of the BP ‘firemen’. Sentence (339b) does of course admit the existential reading which sentence (339a) lacks, namely “there is a firemen who is tall.” Let me thus make sure that my proposal for why the BPS of sentence (339a) lacks the existential reading does *not* rule out sentence (339b) too.

- (339) a. Firemen are tall. * \exists -reading
 b. Some fireman is tall. \exists -reading

The idea is that the contrast in (339) is analogous to the contrast in (99), repeated once more in (340). The crucial difference between the two sentences (340) is the following: the indefinite of the odd sentence (340a) has narrow scope with respect to the universal quantifier over days; by contrast, the indefinite of the fine sentence (340b) has wide scope.

- (340) It is known that the same guy won on every day:
 a. #On every day, a/some fireman won.
 b. A/Some fireman won on every day.

Analogously, the crucial difference between the two sentences (339) is the following: as noted above, the existential BPS of sentence (339a) can only have narrow scope with respect to the generic operator that binds the time argument of the i-predicate ‘tall’; this is not the case for the indefinite of sentence (339b): its LF (341a) yields truth conditions (341b), namely ψ in (334), where existential quantification over firemen has scope over the generic operator.

- (341) a. [[Some fireman] [λx [GEN [x is tall]]]].
 b. $\psi = \lambda w . \exists_x [[\text{fireman}]^w(x) \wedge \text{GEN}_t [\text{in}^w(x, t)]] [[\text{tall}]^w(x, t)]]$.

Let me now show that the blind strengthened meaning of ψ in (341b) is not a contradiction given common knowledge. By the assumption (110) that ‘some fireman’ and ‘the fireman such and such’ are Horn-mates, each ψ_p in (332) is a scalar alternative of ψ in (341b). Note that in the case of ψ in (341), contrary to the case of φ in (330), it is not possible to exclude all such ψ_p ’s, since $\bigwedge_p \neg \psi_p$ is logically equivalent to ψ , as noted in (335). As a matter of fact, by repeating the reasoning illustrated immediately below (106) for the case of sentence (104a), it turns out that none of these alternatives ψ_p ’s in (332) is excludable given ψ in (341). Thus, the blind strengthened meaning of ψ cannot be a contradiction given common knowledge, and the attested existential reading of the indefinite in (339b) is thus sheltered from the MH (97).

Second remark: ‘there’-construction After Milsark (1977), it is well known that i-predicates cannot appear in the coda of the ‘there’-construction contrary to s-predicates, as illustrated by the contrast between (342) and (343).

- (342) a. There are many children sick.
 b. There were people sick.
 c. There were people drunk.
 d. There were doors open.
 e. There were instect that I couldn’t identify on the windowsill.
 f. There were three cars rolling down the hill.
 g. There was a woman running along the beach.
- (343) a. ?There are firemen tall.
 b. ?There were people intelligent.
 c. ?There were doors wooden.
 d. ?There were people doctors.
 e. ?There was a surgeon a happy woman.
 f. ?There was a tycoon that Lynne met on an expedition owning this company.

g. ?There was a student who ordinarily sits in the front row knowing the answer.

My take on this contrast is that the ‘there’-construction forces the indefinite or the bare plural to have narrow scope with respect to the operator that binds the time argument of the i-predicate, either the existential or the generic operator. In other words, sentence (343a) can only have one of the two truth conditions in (344).

$$(344) \quad \lambda w. \left\{ \begin{array}{l} \text{GEN}_t \\ \exists_t \end{array} \right\} [C^w(t)][\exists_x [[\text{firemen}]^w(x) \wedge [[\text{tall}]^w(x, t)]]].$$

These truth conditions are precisely the same truth conditions already considered in (331) and (330), and are therefore ruled out by the MH (97).

3.4 Word order and extraction in German

Sentences (345), from Diesing (1992), contain the BPS ‘Feuerwehrmänner’ (‘firemen’) at the right of the particles ‘ja doch’ (‘indeed’). Sentence (345a) contains the s-predicate ‘verfügbar’ (‘available’), and is fine. Sentence (345b) contains the i-predicate ‘intelligent’, and is bad. The deviance of sentence (345b) is due precisely to the mutual position of ‘ja doch’ and the BPS: once the order is reversed, sentence (345b) turns fully acceptable.

- (345) a. ... weil ja doch Feuerwehrmänner verfügbar sind.
 ... since PARTs firemen available are
- b. # ... weil ja doch Feuerwehrmänner intelligent sind.
 ... since PARTs firemen intelligent are

More examples of the fact that BPSs of i-predicates cannot sit at the right of the particles ‘ja doch’ or of the adverb ‘angeblich’ (‘allegedly’) are provided in (346), from Diesing (1992) and Jäger (2001). Thus, the difference between s- and i-predicates has consequences for German word order: subjects of i-predicates of German subordinate clauses cannot occur to the right of certain particles (such as ‘ja doch’) or certain adverbs (such as ‘angeblich’), contrary to the case of subjects of s-predicates.

- (346) a. # ... weil ja doch Professoren intelligent sind.
 ... since PARTs professors intelligent are
- b. # ... weil ja doch Skorpione giftig sind.
 ... since PARTs scorpions poisonous are.
- c. # ... weil ja doch Wolfshunde Deutsch können.
 ... since PARTs German shepherds German know.
- d. # ... weil angeblich Feuerwehrmänner selbstlos sind.
 ... since allegedly firemen altruistic are

Sentence (347a) illustrates the so-called German SPLIT-TOPIC construction: a subject of the form ‘viele NP’ (‘many NP’) is split by moving the NP to some topic position at the left edge of the clause while stranding ‘viele’. Diesing (1992) notes that split-topic is not available for subjects of i-predicates, as shown by (347b).

- (347) a. Feuerwehrmänner sind viele verfügbar.
 Firemen are many available
 ‘As for firemen, many are available’.
- b. # Feuerwehrmänner sind viele intelligent.
 Firemen are many intelligent
 ‘As for firemen, many are intelligent’.

More examples of the fact that split-topic is not possible with subjects of i-predicates are provided in (348), again from Diesing (1992). Thus, the difference between s- and i-predicates has consequences for extraction in German.

- (348) a. # Wildschweine sind viele intelligent.
 wild boars are many intelligent
 ‘As for wild boars, many are intelligent’
- b. # Schuhe sind viele wasserdicht.
 shoes are many waterproof
 ‘As for shoes, many are waterproof’.
- c. # Linguisten wissen das viele.
 linguists know this many
 ‘As for linguists, many know this’.

In this section, I argue that these contrasts follow from the same piece of reasoning presented in section 3.3 to account for the lack of the existential reading of BPSs of *i*-predicates. The logic of my proposal is analogous to that of Chierchia (1994).

3.4.1 Existing accounts

Diesing argues that both the contrast w.r.t. word order illustrated in (345) and the contrast w.r.t. split-topic illustrated in (347) follow from her assumption (293) — a simplified formulation of which is repeated in (349) — together with plausible specific assumptions on the grammatical phenomena considered. Let me spell out the details of Diesing’s account.

- (349) a. Subjects of *s*-predicates are base generated within VP, in [Spec, VP];
 b. Subjects of *i*-predicates are base generated outside of VP, in [Spec, IP].

Assume that the generalization (350) holds. This assumption (350) entails that subjects at the right of ‘ja doch’ or ‘angeblich’ sit inside VP (plausibly, in [Spec, VP]) while subjects at the left of ‘ja doch’ or ‘angeblich’ sit outside of VP (plausibly, in [Spec, IP]). The contrast w.r.t. word order illustrated in (345) thus immediately follows from Diesing’s crucial assumption (349): “if, as I have claimed, the distinction between *s*- and *i*-predicates is a syntactic distinction that restricts the position of the subject in *i*-predicates but not in *s*-predicates, then the facts in [(345)] are not unexpected. The [sentence (345b)] is expected to be less good, since the subject of an *i*-predicate is base generated in [Spec, IP], the outer position, and has no option of lowering into VP.”

- (350) Certain particles (such as ‘ja doch’) and certain adverbs (such as ‘angeblich’) mark the left edge of VP in German.

Assume that split-topic constructions are derived accordingly to assumption (351), from Riemsdijk (1989). As Diesing points out, this assumption (351) states that the “split-topic construction [is] sensitive to the position of a subject from which extraction occurs. If the subject is VP-internal, extraction is possible. If the subject is VP-external, extraction is not allowed.” The contrast w.r.t. split-topic illustrated in (347) thus immediately follows from Diesing’s crucial assumption (349): “it is predicted that extraction should be possible from the subjects of *s*-predicates, since these subjects have the option of appearing in the VP-internal position. On the other hand, *i*-predicates should disallow extraction from the subject, since they do not permit the option of having the subject in [Spec, VP].”

- (351) The split-topic construction is derived as follows: the subject is base generated as ‘viele NP’ in [Spec, VP]; the surface form is derived by movement (or “extraction”) of NP into some “topic” position at the left edge of the sentence, while the determiner ‘viele’ is stranded behind in situ.

Jäger (2001) provides the following very interesting argument against Diesing’s account for the word order contrast in (345). Diesing’s account rests on the crucial syntactic assumption (349), that draws a distinction between subjects of *s*- and *i*-predicates. This assumption does not distinguish between subjects with different quantificational force. Thus, Diesing’s account for the word order contrast in (345) predicts that the contrast should arise not only with BPs but also with any other type of

subject. As Jäger points out, this prediction is *not* borne out. He notes for instance that universally quantified subjects behave differently than BPSs: as we just saw, BPSs of i-predicates prefer to sit at the left of 'ja doch'; by contrast, "quantifiers headed by 'alle' ('all') or 'jeder' ('every') generally occur preferably to the right of the particle [...], no matter what sort of predicate we take." For instance, sentence (352) is perfectly fine, despite the fact that the subject 'alle Studenten' ('every student') of the i-predicate 'können' ('to know') occurs at the right of the particle 'ja'.³³

- (352) ... weil ja alle Studenten Englisch können
 ... because PART all students English know.

3.4.2 An account based on blind and mandatory mismatching implicatures

In this subsection, I present an account for the two contrasts (345) and (347) that retains Diesing's two specific assumptions (350) and (351) but dispenses with her problematic assumption (349).

Preliminaries Let me start with sentence (345a) repeated below in (353), with the s-predicate 'verfügbar' ('available') and the BPS at the right of the particles 'ja doch'. As Diesing points out, the BPS 'Professoren' ('professors') of sentence (353) *only* admits the existential reading, which is compatible with the predicate being construed either as episodic as in (353a), or as habitual as in (353b). The generic reading of the BPS as in (353c) is unattested, and only available for sentence (354) with the inverted word order, namely the BPS at the left of the particles 'ja doch'.

- (353) ... weil ja doch Professoren verfügbar sind.
 ... since indeed professors available are
 a. = There are professors available.
 b. = Usually, there are professors available.
 c. ≠ Professors are (in general) available.
- (354) ... weil Professoren ja doch verfügbar sind.
 ... since professors indeed available are
 a. ≠ There are professors available.
 b. ≠ Usually, there are professors available.
 c. = Professors are (in general) available.

Thus, sentence (353) only admits the two LFs (355a) and (356a), which yield the two truth conditions (355b) and (356b) respectively, which correctly capture the two attested readings (353a) and (353b). Note crucially that the the existential BPS has narrow scope with respect to the operator that binds the time argument of the predicate. No other LF is available for sentence (353), since it has no any other reading, in particular it does not admit the generic reading for its BPS.³⁴

³³ Jäger furthermore notes that sentence (352) with the subject 'alle Studenten' ('all students') at the left of 'ja' is less than perfect for a few speakers. I have nothing to say about this fact.

³⁴ For the sake of explicitness, let me digress to ask the following question: why is it that the BPS of an s-predicate at the right of 'ja doch' can only be interpreted existentially, as in (353)? Or, more explicitly: why is it that the BPS of an s-predicate at the left of 'ja doch' can only be interpreted generically, as in (354)? As Diesing herself suggests, the solution must hinge on the fact that scope relationships in German are fixed at S-structure, as stated in (i).

- (i) In German, a scope bearing element α can have scope over a scope bearing element β at LF iff α c-commands β at S-structure.

Assumption (i) can be illustrated by means of the following example (ii), from Jäger (2001): contrary to its English translation, the German sentence (ii) can have only the reading (iia), which corresponds to the surface scope configuration 'sometimes' > 'all'; but it cannot have the reading (iib), which corresponds to the opposite scope configuration 'all' > 'sometimes'.

- (ii) ... weil manchmal alle Studenten die Antwort wissen.
 ... since sometimes all students the answer know.
 a. = Sometimes, all the students know the answer.
 b. ≠ Every student knows sometimes the answer.

- (355) LF and truth-conditions for the reading (353a):
- [ja doch [GEN_t [_{VP} firemen(*x*) available(*t*)]]].
 - $\lambda w . \text{GEN}_t [C^w(t)] [\exists_x [[\text{firemen}]^w(x) \wedge [\text{available}]^w(x, t)]]$.
- (356) LF and truth-conditions for the reading (353b):
- [ja doch [\exists_t [_{VP} firemen(*x*) available(*t*)]]].
 - $\lambda w . \exists_t [C^w(t)] [\exists_x [[\text{firemen}]^w(x) \wedge [\text{available}]^w(x, t)]]$.

Account for the word order contrast in (345) Under the assumption that there is no relevant grammatical difference between *s*- and *i*-predicates, exactly the same LFs which are available for sentence (345a) with the *s*-predicate ‘verfügbar’ (‘available’) should also be available for sentence (345b) with the *i*-predicate ‘intelligent’. Thus, this latter sentence (345b) should have the two LFs in (357a) and (358a), and no others. The corresponding truth conditions are φ and φ' in (357b) and (358b). These truth conditions φ and φ' in (357b) and (358b) are identical to the truth conditions φ and φ' in (330b) and (331b) considered in section 3.3 for the unattested existential reading of the BPS ‘firemen’ of the basic sentence ‘Firemen are tall’. Hence, the blind strengthened meaning of the truth conditions φ and φ' in (357b) and (358b) contradicts the piece of common knowledge (227). In conclusion, all the LFs corresponding to sentence (345b) lead to truth conditions ruled out by the by MH (97) and the deviance of sentence (345b) is thus predicted.

- (357) a. [ja doch [GEN_t [_{VP} firemen(*x*) intelligent(*t*)]]].
- b. $\varphi \doteq \text{GEN}_t [C^w(t)] [\exists_x [[\text{firemen}]^w(x) \wedge [\text{intelligent}]^w(x, t)]]$.
- (358) a. [ja doch [\exists_t [_{VP} firemen(*x*) intelligent(*t*)]]].
- b. $\varphi' \doteq \exists_t [C^w(t)] [\exists_x [[\text{firemen}]^w(x) \wedge [\text{intelligent}]^w(x, t)]]$.

My account for the contrast in (345) rests on specific assumptions on BPSs, namely that a BPS such as ‘firemen’ contributes an existential quantifier to the truth conditions and furthermore is a Horn-mate of the definite description ‘the fireman such and such’, by assumption (333). Thus my account leaves open the possibility that other types of subjects of *i*-predicates might differ from BPSs in being perfectly fine at the right of ‘ja doch’. In particular, since universal quantifiers are maximal in their Horn-scale, then blind strengthening is vacuous and thus cannot trigger any mismatch with common knowledge. Thus, nothing in my proposal rules out a universally quantified subject at the right of ‘ja doch’. In conclusion, my account is well compatible with Jäger’s observation (352), repeated in (359), that universally quantified subjects are fine at the right of the particles ‘ja doch’.³⁵

- (359) ... weil ja alle Studenten Englisch können
... because PART all students English know.

Account for the split-topic contrast in (347) By assumption (351), split-topic is obtained as follows: the subject ‘viele NP’ is base generated within [Spec, VP] and then NP is extracted to some topic position. The LF scope of the subject could in principle be either its base position within VP or its landing position outside VP. It turns out that the attested LF scope is the former. Here is a way to see that. In the case of sentence (360), the Q-adverb ‘gewöhnlich’ (‘usually’) intervenes between the base position and the landing position of the NP ‘Freunde’ (‘friends’). If the LF scope of the subject

The patterns in (353) and (354) follow straightforwardly from assumption (ii). The BPS of sentence (353) is at the right of ‘ja doch’; by assumption (350), the BPS is thus inside VP at S-structure; by assumption (ii), it must remain there at LF; since Diesing assumes that variables introduced by material in the VP can only be bound by default existential closure, it then follows that this BPS can only receive the existential reading. The BPS of sentence (354) is instead at the left of ‘ja doch’; by assumption (350), it is thus outside of VP at S-structure; by assumption (ii), it must remain there at LF; since Diesing assumes that variables introduced by material outside of VP cannot be bound by default existential closure and must instead be bound by a generic operator, it thus follows that this BPS can only receive the generic reading.

³⁵I am being sloppy here. My proposal predicts that sentence (359) should have something like the truth conditions in (i), where \forall_x is a short hand for ‘for all students’.

(i) $\text{GEN}_t [C(t)] [\forall_x [[\text{know-English}](x, t)]]$

The problem is that I do not quite know what should the restrictor *C* of the generic operator look like.

'viele Freunde' ('many friends') were the landing position of extraction, then we would expect the reading in (360b) to be attested, which is not, as pointed out to me by Patrick Grosz (p.c.). The only attested reading is (360a), that corresponds to the subject 'viele Freunde' having LF scope in the base position within VP.

- (360) dass Freunde ihn gewöhnlich viele übers Ohr hauen.
 that friends him usually many over.the ear beat.
 a. = it is usually the case that many friends rip him off.
 b. ≠ there are many friends who usually rip him off.

Thus, sentence (347b) must correspond to one of the two truth conditions in (361). These truth conditions are identical to those in (357b) or (358b), only with \exists_x replaced by MANY_x . If assumption (333), that 'firemen' and 'the fireman such and such' are Horn-mates, is extended to the assumption that 'many firemen' and 'the fireman such and such' are Horn-mates too, then these truth conditions (361) are ruled out again by the fact that they yield a contextually contradictory strengthened meaning, just as the truth conditions (357b) and (358b).³⁶

- (361) a. $\varphi \doteq \text{GEN}_t [C^w(t)] [\text{MANY}_x [[\text{firemen}]^w(x) \wedge [\text{intelligent}]^w(x, t)]]$.
 b. $\varphi' \doteq \exists_t [C^w(t)] [\text{MANY}_x [[\text{firemen}]^w(x) \wedge [\text{intelligent}]^w(x, t)]]$.

First remark: definites Jäger also reports that definite subjects behave as BPSs, namely that both types of subjects cannot occur at the right of 'ja'. For instance, sentence (362b) is deviant, precisely because of the definite subject 'der Präsident' ('the president') of the *i*-predicate 'intelligent' at the right of 'ja'.

- (362) a. ... weil der Präsident ja intelligent ist.
 ... since the president PART intelligent is.
 b. * ... weil ja der Präsident intelligent ist.
 ... since PART the president intelligent is.

The deviance of sentence (362b) is problematic for my proposal; here is why. The LF of sentence (362b) is (363a) and the corresponding truth conditions are φ in (363b). If the restrictor $C^w(t)$ is set by context to be the life span $\text{in}^w([\text{the president}]^w, t)$ of the president, then there is no way that strengthening φ can lead to a contradiction given common knowledge.

- (363) a. $[\text{ja}]_{[\text{VP}]} [[\text{der Präsident}] [\text{intelligent ist}]]$.
 b. $\varphi = \text{GEN}_t [C^w(t)] [[\text{intelligent}]^w([\text{the president}]^w, t)]$.

The problem can be restated as follows: since the definite subject is scopally inert, its position has no truth conditional consequences and thus my semantically-based approach predicts that it shouldn't matter where the definite sits, contrary to what shown by the contrast in (362). I have to leave the issue open for the time being.

³⁶Diesing (1992) notes that the *s-i*-predicates distinction also has consequences for the German 'was für'-split construction. The German DP 'was für N' means something like 'what kind of N'. As shown in (ib), it is possible to move the *wh*-element 'was' to some higher position (plausibly [Spec, CP]) stranding behind the rest of the DP. This construction is called 'WAS-FÜR' SPLIT. Diesing notes that 'was-für' split of a subject DP depends on the predicate: as shown in (ib), it is possible in the case of the subject of the *s*-predicate 'verfügbar' ('available'); as shown in (ia), it is impossible in the case of the subject of the *i*-predicate 'intelligent'.

- (i) a. * Was sind für Leguane intelligent?
 what are for iguanas intelligent
 'What kind of iguanas are intelligent?'
 b. Was sind für Leguane verfügbar?
 what are for iguanas available
 'What kind of iguanas are available?'

I have no proposal to make concerning the contrast in (i).

Second remark: German negative BPSs Kratzer (1995, p. 144-145) notes that s- and i-predicates in German behave differently with respect to plural negative subjects of the form ‘keine NP_{pl}’, as illustrated by the following contrast: the sentences in (364) have s-predicates, and they are fine; the sentences in (365) have i-predicates, and they are bad.

- (364) a. ... weil uns keine Freunde helfen.
 ... since us no friends help.
 ‘... since no friends are helping us’.
- b. ... weil hier keine Fliederbäume wachsen.
 ... since here no lilacs grow.
 ‘... since no lilacs are growing here’.
- (365) a. * ... weil keine Ärzte altruistisch sind.
 ... since no physicians altruistic are.
 ‘... since no physicians are altruistic’.
- b. * ... weil das keine Kandidaten wissen.
 ... since this no candidates know.
 ‘... since no candidates know this’.

Kratzer argues that the contrast between (364) and (365) follows from assumption (349), concerning the different base positions of subjects of i- and s-predicates, together with the auxiliary assumption (366). A sentence such as (364a) is fine because it admits the following derivation: according to assumption (349), the BPS ‘Freunde’ (‘friends’) is base generated in [Spec, VP]; according to assumption (366b), ‘nicht’ is base generated next to it, say adjoined to VP; according to assumption (366a), ‘nicht’ and the existential operator merge as ‘keine’. No such derivation is possible for a sentence such as (365a): according to assumption (366a), the phrase ‘keine Ärzte’ (‘no doctors’) can only be generated as a result of the adverb ‘nicht’ (‘no’) and the (existential closure operator associated with the) BP ‘Ärzte’ (‘doctor’) being base generated next to each other; but this can never happen, given that ‘nicht’ is base generated at the edge of VP by assumption (366b) while the BP is base generated outside of VP by assumption (349).³⁷

- (366) a. At D-structure and at LF, a quantifier phrases such as ‘keine NP’ (‘no NP’) consists “of the negation adverb ‘nicht’ (‘not’) and the bare plural NP. The bare plural has to receive existential force through existential closure. ‘Keine NP’, then, involves two operators: negation and the existential closure operator.” There is a phonological rule which mandatorily realizes ‘nicht’ and an adjacent (plural) existential closure operator as ‘kein—’.
- b. “Just like its English counterpart ‘not’, ‘nicht’ is base generated somewhere between the subject and the VP, that is, in the usual position for sentence adverbs”.

Thus, it looks like Diesing’s assumption (349) on the base position of subjects of s- and i-predicates offers a straightforward account of the contrast between (364) and (365). And I see no way of recasting this account within the purely semantic framework I have defended so far. Yet, Sauerland

³⁷Evidence for assumption (366a) is provided by sentence (i). If ‘keine Beispiele’ were a normal quantifier, we would expect sentence (i) to have either the reading in (ia) or that in (ib), depending on the relative scope of this quantifier with the modal. However, sentence (i) has neither of these readings, but rather the one in (ic). This last reading is predicted by assumption (366a): if ‘keine’ is split into ‘nicht’ and an existential operator at LF, then a third operator can intervene between the two, as in (37c).

- (i) ... weil keine Beispiele bekannt sein müssen.
 ... since no examples known be must.
- a. $\neq \text{NO}_x \left[\left[\text{example} \right] (x) \right] \left[\left[\text{known} \right] (x) \right]$.
- b. $\neq \left[\left[\text{example} \right] (x) \right] \left[\left[\text{known} \right] (x) \right]$.
- c. $= \text{NOT} \left[\left[\text{known} \right] (x) \right] \left[\left[\text{example} \right] (x) \right]$.

See also Sauerland (2000) and Penka and Zeijlstra (2005) for further discussion of this split-scope reading (also, note that Penka and Zeijlstra report slightly different judgments).

(2007) has very recently pointed out some new interesting facts concerning the distribution of 'keine NP_{pl}', that cast doubts on Kratzer's account. He considers predicates such as 'to be related by blood' and 'to be genetically identical'. These predicates behave as standard i-predicates: their BPSs only admit the generic reading, as shown in (367a); and they resist 'there'-construction and 'when'-clauses, as shown in (367b) and (367c).

- (367) a. Royals are related by blood.
 b. #There are royals related by blood.
 c. #When John is related by blood to Mary, he is her brother.

Yet, contrary to the i-predicates in sentences (365), these i-predicates can occur with plural negative indefinites, as illustrated by the fact that sentences (368) are fine, both with and without 'zwei' ('two'). Sauerland points out that the felicity of (368) poses a problem for Kratzer's account of the distribution of German plural negative quantifiers. In fact, in order to account for the lack of the existential reading of the BPS 'royals' of sentence (367a), Diesing and Kratzer would have to assume that subjects of the predicate 'miteinander blutsverwandt sein' ('to be related by blood one with the other') are base generated directly in [Spec, IP]. But assumptions (366) would then incorrectly predict that sentence (368a) cannot be generated.

- (368) a. ... weil keine ('zwei') Linguisten miteinander blutsverwandt sein.
 ... since no (two) linguists with-one-another blood-related are.
 'Since no linguists are related by blood'.
 b. ... weil keine (zwei) Zebrafische genetisch identisch sind.
 ... since no (two) zebra-fish genetically identical are.
 'Since no (two) zebra fish are genetically identical'.

The crucial difference between the i-predicates in (365) and those in (368) is that the former are inherently distributive while the latter are inherently collective. The distribution of German plural negative quantifiers just described thus looks similar to the distribution of French plural 'des'-indefinites with generic meaning, as illustrated in (284)-(283) from Dobrovie-Sorin (2003).

3.5 BPSs embedded under a universal operator

Fox (1995) observes that the existential reading of the BPS of an i-predicate becomes available when the BPS is in the scope of a universal operator, such as a universal object. As expected, the BPS 'Jewish women' of the i-predicate 'to be related to Chomsky' of sentence (369a) does not admit the existential reading. In sentence (369b), the definite 'Chomsky' has been replaced by the universal object 'every Jewish man' and surprisingly the BPS 'Jewish women' does admit the existential reading, as soon as the universal object is given wide scope, i.e. the sentence can mean that "for every Jewish man *there are* Jewish women who are related to him."³⁸

- (369) a. Jewish women are related to Chomsky. #∃-BPS
 b. Jewish women are related to every Jewish man. ∃-BPS
 (370) a. MIT students know Chinese. #∃-BPS
 b. MIT students know almost every foreign language. ∃-BPS

³⁸I wonder whether the contrast just pointed out holds also for the i-predicates 'love' and 'hate'.

- (i) a. MIT professors love Chomsky.
 b. MIT professors love every Jewish linguist.
 (ii) a. MIT professors hate Chomsky.
 b. MIT professors hate every Jewish linguist.

It is not completely clear to me whether there is a contrast between the (a) and (b) sentences in (i) and (ii).

As Fox points out, the existential interpretation of the BPSs in (369b) and (370b) is only possible when the subject has narrow scope with respect to the universal object, as shown by the fact that the existential reading disappears in the variants in (371), where the subject binds a pronoun within the object and is therefore forced to have wide scope.

- (371) a. Jewish women are related to every Jewish man they know.
 b. MIT students know almost every foreign language they've heard of.

In this section, I'll argue that the contrast illustrated in (369)-(370) follows from the proposal developed so far: the reasoning developed in section 3.3 to rule out the existential reading of BPSs of *i*-predicates fails when BPSs are embedded under a universal quantifier (over individuals). This should be no surprise, for the following reason. My account for the lack of the existential reading of BPSs of *i*-predicates is identical to the account proposed in section 2.2 for why sentence (99a), repeated in (372a), sounds odd. But sentence (99a) turns acceptable when it is embedded under a further universal operator, as shown by sentence (101b), repeated in (372b).

- (372) *Context*: a competition lasted for five days, Monday through Friday; each day, three challenges are held: swimming, running and jumping; both John and Bill know that the same guy *x* won the swimming competition on all five days, the same guy *y* won the running competition on all five days and the same guy *z* won the jumping competition on all five days:

	M	T	W	T	F
<i>swimming</i> :	<i>x</i>	<i>x</i>	<i>x</i>	<i>x</i>	<i>x</i>
<i>running</i> :	<i>y</i>	<i>y</i>	<i>y</i>	<i>y</i>	<i>y</i>
<i>jumping</i> :	<i>z</i>	<i>z</i>	<i>z</i>	<i>z</i>	<i>z</i>

John wants to know more about these amazing guys, and thus asks Bill for more information; Bill replies as follows:

- a. ?Every day, a/some fireman won the running competition.
 b. Every day, for every competition, a/some fireman won.

Thus, my account for Fox's effect, that embedding under a universal operator rescues the existential reading of BPSs of *i*-predicates, is identical to the account suggested in section 2.2 for why embedding under a universal operator rescues sentence (372b). To stress the analogy between the two accounts for the contrasts in (372) and (369), Table 3.6 summarizes the main steps of the two accounts, one next to the other.

3.5.1 Existing accounts

Fox (1995) suggests an account for the contrast in (369) based on three ingredients: the first ingredient is Diesing's assumption (326) that existential BPSs come about through the DEO and that the DEO has VP scope; the second ingredient is Kratzer's assumption (236) that *i*-predicates lack a time argument; the third assumption is Fox's SCOPE ECONOMY (373), that roughly bans LF-movement which does not have truth conditional consequences. Let me present the details of his account.

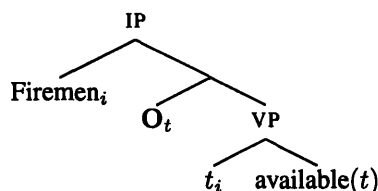
- (373) LF-movement of a quantifier Q_1 which does not cross at least one quantifier Q_2 such that Q_1 and Q_2 do not commute is ungrammatical.

The case of BPSs of *s*-predicates is illustrated in (374). The subject is base generated in [Spec, VP] and moved to [Spec, IP]. The time argument of the *s*-predicate is bound by a tense operator O_t which sits between the surface position [Spec, IP] of the BPS and its base position [Spec, VP]. By Diesing's assumption (326), the BPS needs to be reconstructed into VP in order to get bound by the DEO and thus be interpreted existentially. Since the BPS and the tense operator don't commute, reconstruction of the BPS is allowed by Scope Economy (373). The existential reading of BPSs of *s*-predicates is thus accounted for.

Every day for every competition, a fireman won.	Jewish women are related to every Jewish man.																																																
(1) LF AND TRUTH CONDITIONS																																																	
<p style="text-align: center;">φ</p> <p style="text-align: center;">[every competition]_y every day_t a fireman_x won</p> <p style="text-align: center;">$\varphi = \lambda w . \forall_y [\forall_t [\exists_x [[\text{won}]^w(x, y, t)]]]$</p>	<p style="text-align: center;">φ</p> <p style="text-align: center;">[every man]_y GEN_t women_x related</p> <p style="text-align: center;">$\varphi = \lambda w . \forall_y [\text{GEN}_t [\exists_x [[\text{related}]^w(x, y, t)]]]$</p>																																																
(2) SET OF SCALAR ALTERNATIVES																																																	
$\text{Alt}(\varphi) = \{\psi_P \mid P \in \mathcal{D}_{(e,t)}^{\mathcal{W}}\}$, where ψ_P is the LF with the following shape and truth conditions																																																	
<p style="text-align: center;">ψ_P</p> <p style="text-align: center;">[every competition]_y every day_t the fireman P won</p> <p style="text-align: center;">$\psi_P = \lambda w . \forall_y [\forall_t [[\text{won}]^w(d_P^w, y, t)]]]$</p>	<p style="text-align: center;">ψ_P</p> <p style="text-align: center;">[every man]_y GEN_t the woman P related</p> <p style="text-align: center;">$\psi_P = \lambda w . \forall_y [\text{GEN}_t [[\text{related}]^w(d_P^w, y, t)]]]$</p>																																																
(3) A USEFUL RESTATEMENT OF THE NEGATION OF ALL THE ALTERNATIVES																																																	
$\bigwedge_P \neg \psi_P = \neg \psi$, where ψ are the following truth conditions																																																	
$\psi \doteq \lambda w . \exists_x [\forall_y [\forall_t [[\text{win}]^w(x, y, t)]]]$	$\psi \doteq \lambda w . \exists_x [\forall_y [\text{GEN}_t [[\text{related}]^w(x, y, t)]]]$																																																
(4) SET OF EXCLUDABLE ALTERNATIVES																																																	
$\text{Excl}(\varphi) = \{\psi_P \mid P \in \mathcal{D}_{(e,t)}^{\mathcal{W}}\}$, since $\varphi \wedge \neg \psi$ is true e.g. in the following world:																																																	
<table border="1" style="border-collapse: collapse; width: 100%;"> <thead> <tr> <th></th> <th>M</th> <th>T</th> <th>W</th> <th>T</th> <th>F</th> </tr> </thead> <tbody> <tr> <td>swimming:</td> <td><i>x</i></td> <td><i>x</i></td> <td><i>x</i></td> <td><i>x</i></td> <td><i>x</i></td> </tr> <tr> <td>running:</td> <td><i>y</i></td> <td><i>y</i></td> <td><i>y</i></td> <td><i>y</i></td> <td><i>y</i></td> </tr> <tr> <td>jumping:</td> <td><i>z</i></td> <td><i>z</i></td> <td><i>z</i></td> <td><i>z</i></td> <td><i>z</i></td> </tr> </tbody> </table>		M	T	W	T	F	swimming:	<i>x</i>	<i>x</i>	<i>x</i>	<i>x</i>	<i>x</i>	running:	<i>y</i>	<i>y</i>	<i>y</i>	<i>y</i>	<i>y</i>	jumping:	<i>z</i>	<i>z</i>	<i>z</i>	<i>z</i>	<i>z</i>	<table border="1" style="border-collapse: collapse; width: 100%;"> <thead> <tr> <th></th> <th><i>t</i>₁</th> <th><i>t</i>₂</th> <th><i>t</i>₃</th> <th><i>t</i>₄</th> <th><i>t</i>₅</th> </tr> </thead> <tbody> <tr> <td>John:</td> <td>Sue</td> <td>Sue</td> <td>Sue</td> <td>Sue</td> <td></td> </tr> <tr> <td>Bill:</td> <td></td> <td>Ann</td> <td>Ann</td> <td>Ann</td> <td>Ann</td> </tr> <tr> <td>Tom:</td> <td></td> <td></td> <td>Mary</td> <td>Mary</td> <td>Mary</td> </tr> </tbody> </table>		<i>t</i> ₁	<i>t</i> ₂	<i>t</i> ₃	<i>t</i> ₄	<i>t</i> ₅	John:	Sue	Sue	Sue	Sue		Bill:		Ann	Ann	Ann	Ann	Tom:			Mary	Mary	Mary
	M	T	W	T	F																																												
swimming:	<i>x</i>	<i>x</i>	<i>x</i>	<i>x</i>	<i>x</i>																																												
running:	<i>y</i>	<i>y</i>	<i>y</i>	<i>y</i>	<i>y</i>																																												
jumping:	<i>z</i>	<i>z</i>	<i>z</i>	<i>z</i>	<i>z</i>																																												
	<i>t</i> ₁	<i>t</i> ₂	<i>t</i> ₃	<i>t</i> ₄	<i>t</i> ₅																																												
John:	Sue	Sue	Sue	Sue																																													
Bill:		Ann	Ann	Ann	Ann																																												
Tom:			Mary	Mary	Mary																																												
(5) CONCLUSION																																																	
The sentences denote the proposition $\text{EXH}(\varphi) = \varphi \wedge \neg \psi$, which is not a contradiction w.r.t. common knowledge, since the possible world considered in step (4) is in \mathcal{W}_{ck} .																																																	

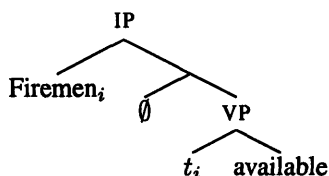
Table 3.6: Parallelism between the account for the contrast in (372) presented in section 2.2 and the account for the contrast in (369) presented in this section.

(374)



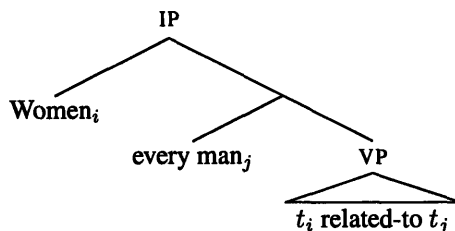
The case of BPSs of *i*-predicates is illustrated in (375). Again, the subject is base generated in [Spec, VP] and moved to [Spec, IP], because that is always the case, both for *s*- and *i*-predicates. By Kratzer's assumption (236), *i*-predicates lack a time argument and thus there is no tense operator above VP. Again by Diesing's assumption (326), the BPS would need to be reconstructed into VP in order to get bound by the DEO and thus be interpreted existentially. But in this case lowering of the BPS into VP is blocked by Scope Economy (373), because there is no operator to cross. The lack of the existential reading of BPSs of *i*-predicates is thus accounted for.

(375)



Finally, the case of BPSs of *i*-predicates with wide scope universal objects is illustrated in (376). The universal object has been QRed at the left edge of VP, to resolve a type mismatch. Again, the subject is base generated in [Spec, VP] and moved to [Spec, IP]. Again, the *i*-predicate lacks a time argument and thus there is no tense operator above VP. Again, the BPS needs to be reconstructed into VP in order to get bound by the DEO and thus be interpreted existentially. In this case, lowering of the BPS into VP is not blocked by Scope Economy (373), because movement of the BPS crosses the object adjoined to [Spec, VP] which does not commute with the BPS.

(376)



In conclusion, Fox offers an account for the restriction (287) on the existential reading of BPSs of *i*-predicates that does without Diesing's problematic assumption (293), concerning the different base positions of subjects of *s*- and *i*-predicates. Furthermore, Fox's account predicts the effect in (369)-(370) of embedding underneath a universal object. To evaluate Fox's account, let me step back and look at the overall picture on BPSs of *i*-predicates. In sections 3.3 and 3.4, I have considered two restrictions on BPSs of *i*-predicates, recalled in (377a) and (377b), respectively. Both Diesing's and mine accounts for the restriction (377a) on the interpretation of BPSs of *i*-predicates straightforwardly extends to the restriction (377b) on German word order, through the auxiliary assumption (350). This is not true for Fox's proposal: since his account for the restriction (377a) on the interpretation of BPSs is based on an economy constraint on Quantifier Lowering, it does not extend to the word order restriction in (377b), where movement does not seem to play any role. The restriction (377b) is thus left unexplained by Fox's proposal.

- (377) a. BPSs of *i*-predicates lack the existential reading;
 b. BPSs of *i*-predicates cannot sit at the right of the particles 'ja doch' in German.

The desirability of a unified analysis for the two restrictions (377a) and (377b) is underscored by their common sensitivity to embedding underneath a universal object. Fox's examples (369)-(370) show the sensitivity of the interpretational restriction (377a) to embedding underneath a universal

object. Let me now show the sensitivity of the word order restriction (377b) to embedding underneath a universal object by means of the contrast in (378). Sentence (378a), with the i-predicate ‘verwandt’ (‘related to’) and the BPS ‘jüdische Frauen’ (‘Jewish women’) at the right of ‘ja doch’ is deviant, as expected given the contrast in (3.4) considered in section 345. Yet, sentence (378b), where the definite object ‘Chomsky’ has been replaced by the universal quantifier ‘jedem jüdischen Mann’ (‘every Jewish man’), sounds fine. The contrast in (378) is thus a replication of Fox’s contrast for the interpretational restriction (377a) to the case of the word order restriction (377b).

- (378) a. #Mit Chomsky sind *ja doch* jüdische Frauen verwandt.
with Chomsky are PARTs Jewish women related
b. Mit jedem Jüdischen Mann sind *ja doch* jüdische Frauen verwandt.
with every Jewish man are PARTs Jewish women related

Let me finally note that the contrast in (378) is also problematic for Diesing’s account for the word order restriction (377b) reviewed in subsection 3.4.1, since her proposal is not sensitive to properties of the embedding environment.

3.5.2 An account based on blind and mandatory mismatching implicatures

In this subsection, I present my account for Fox’s effect of embedding under a universal operator over individuals, and then turn to the case of embedding under a modal.

Account Of course, the existential reading of the BPS of sentence (369a), repeated in (379a), is ruled out in the same way as the existential reading of the BPS of the basic sentence (287b), repeated in (379b), since the reasoning illustrated in section 3.3 can be repeated for the i-predicate ‘related to Chomsky’ in place of ‘tall’.

- (379) a. Jewish women are related to Chomsky.
b. Firemen are tall.

Let’s now see what happens when this same line of reasoning is applied to (369b), with the wide scope universal object ‘every Jewish man’. The LF which corresponds to the existential reading of the BPS of this sentence (369b) is the one in (380): the BPS ‘Jewish women’ has been reconstructed into [Spec, VP]; the universal object ‘every Jewish man’ has been QRed out of VP; the time argument of the i-predicate ‘related’ is bound by the generic operator; the restrictor of the generic operator is a predicate C assigned by the context.

- (380) [every man_y [[GEN_t C] [_{VP} women(x) related-to(y , t)]]].

This LF (380) yields the truth-conditions φ in (381): the DEO binds the variable x introduced by the BPS ‘Jewish women’; the presupposition of the nuclear scope of the generic operator (whatever it is) is added to its restriction C , yielding the predicate \tilde{C} ; the notation \forall_y is a shorthand for the denotation of ‘for every Jewish man’. Thus, I need to run the reasoning in section 3.3 for the case of these truth conditions φ and make sure that they are *not* ruled out.

- (381) $\varphi = \lambda w . \forall_y [\text{GEN}_t [\tilde{C}(y, t)] [\exists_x [[\text{women}]^w(x) \wedge [\text{related}]^w(x, y, t)]]]]$.

By replacing ‘Jewish women’ in (380) with the definite description ‘the Jewish woman such and such’, we get the LF (382a), where I am using ‘the woman P ’ as a short hand for ‘the Jewish woman such and such’. This LF yields the truth conditions in (382b), where I am using d_P^w as a shorthand for [the woman P]^w.

³⁹Note that I am assuming in (381) that the restrictor \tilde{C} depends on the specific Jewish man y . This assumption is needed independently from the issue of i-predicates, for instance in order to get the truth conditions (ib) for sentence (ia) with a universal subject and an habitual s-predicate.

- (i) a. Every friend of mine smokes.
b. $\lambda w . \forall_y [[\text{friend}]^w(y) \rightarrow \text{GEN}_t [C^w(y, t)] [[\text{smoke}]^t(y, t)]]$.

Note that the dependence of the restrictive clause of the generic operator on the individual y might be due to the addition to the restrictive clause of the presupposition of the nuclear scope; I leave the issue open.

- (382) a. [every man_y [[GEN_t C] [the woman *P* related-to(*y*, *t*)]]]
 b. $\psi_P = \lambda w . \forall_y [\text{GEN}_t [\text{C}(y, t) \wedge \text{in}^w(d_P^w, t)] [[\text{related}]^w(d_P^w, y, t)]]$.

Let ψ be as in (383), where \exists_x is a shorthand for the existential quantifier over Jewish women and \forall_y is a shorthand for the universal quantifier over Jewish men. This formula says that “there is a Jewish woman who is related to every Jewish man at all times in the restrictive clause”. Note that ψ only differs from φ in (381) because of the fact that the existential quantifier over Jewish women \exists_x has widest scope. Note furthermore the following crucial difference between (383) and (334): in the case of ψ in (383), the existential quantifier over Jewish women \exists_x has wide scope not only over the generic operator GEN_t but also over the universal quantifier over Jewish men \forall_y .

$$(383) \quad \psi = \lambda w . \exists_x [\forall_y [\text{GEN}_t [\text{C}^w(y, t) \wedge \text{in}^w(x, t)] [[\text{related}]^w(x, y, t)]]]$$

The logical equivalence in (384) holds;⁴⁰ in words: the negation of all the alternatives ψ_P 's in (382b) is equivalent to the negation of ψ in (383), namely to the statement that “there is no Jewish woman who is related to every Jewish man throughout the entire restrictive clause.”

$$(384) \quad \bigwedge_P \neg \psi_P = \neg \psi$$

Let me now argue that the blind strengthened meaning $\text{EXH}(\varphi)$ of our sentence φ in (381) is the one in (385), i.e. that all the alternatives ψ_P 's in (382b) are excludable w.r.t. φ in (381). Given the equivalence in (384), it suffices to show that $\varphi \wedge \neg \psi$ is not a logical contradiction.

$$(385) \quad \text{EXH}(\varphi) = \varphi \wedge \underbrace{\bigwedge_P \neg \psi_P}_{\neg \psi}$$

To this end, consider the possible world w described in (386): there are only two Jewish men a_1 and a_2 ; there are only two Jewish women b_1 and b_2 ; the woman b_1 is related only to the man a_1 and the woman b_2 is related only to the man a_2 . Note that φ is true in such a world w (with a judicious choice of the restrictive clause C), since for every Jewish man there is a Jewish woman related to him. Note furthermore that ψ is false in such a world w , since there is no Jewish woman related to every Jewish man. Thus, $\varphi \wedge \neg \psi$ is true in such a world w and $\text{EXH}(\varphi) = \varphi \wedge \neg \psi$ is not a logical contradiction. Furthermore, the world w in (386) could very well be compatible with the common knowledge \mathcal{W}_{ck} in (227). Hence, the strengthened meaning $\text{EXH}(\varphi)$ of φ in (381b) is *not* a contradiction given common knowledge and the existential reading for the BPS of (369b) is *not* ruled out by the MH.⁴¹

- (386) a. [Jewish man]^w = { a_1, a_2 }.
 b. [Jewish woman]^w = { b_1, b_2 }.
 c. [related-to]^w = {(a_1, b_1), (a_2, b_2)}.

In this section, I have noted that the account for the lack of the existential reading of BPSs of *i*-predicates suggested in section 3.3 predicts the existential reading to be rescued by embedding under a universal object. Fox's contrast (369) showed that this prediction is borne out. In section 3.4, I have suggested that the restriction on German word order illustrated in (345) follows from the same account suggested in section 3.3 for the lack of the existential reading of BPSs of *i*-predicates. By parity of reasoning, my account for German word order thus predicts that embedding under a universal object should rescue the ordering of the BPS at the right of 'ja doch'. This prediction is borne out by the contrast in (378).

⁴⁰The reasoning is identical to that in footnote 16.

⁴¹Of course the problem highlighted in the First remark in section 2.2 holds also in this case. Let me recall the problem. Consider the LFs obtained from the LF in (380) by replacing both 'Jewish women' by 'the Jewish woman such and such' and 'every Jewish man' by 'some Jewish man'. If these LFs were scalar alternatives of (380), then the blind strengthened meaning of the existential reading of the BPS of sentence (369b) would mismatch with common knowledge, and the attested existential reading would be incorrectly ruled out. Crucially, these LFs are *not* scalar alternatives of (381) according to Fox's alternative definition (124) of the set of scalar alternatives.

Remark: embedding under universal modals Fox's contrast shows that the existential reading of a BPS of an *i*-predicate is rescued by embedding under a universal operator. Crucially, not any universal operator does the trick: universal operators over individuals do; but for instance universal modals don't. Despite the fact that the BPS 'firemen' of the *i*-predicate 'tall' is embedded under a universal modal in (387a), it lacks the existential reading, as in the plain case (387b).

- (387) a. It must be the case that firemen are tall. * \exists -BPS
 b. Firemen are tall. * \exists -BPS

Let me thus make sure that my proposal captures this difference between universal quantifiers over individuals and universal modals. The idea is that what's special about modals is that they quantify over worlds and that the scalar item [the fireman P]^{*w*} has world-dependence built into it, since P is a function from worlds into sets of individuals. In order to make this point explicit, I will now quickly repeat for the case of sentence (387a) the piece of reasoning presented in section 3.3 for the case of sentence (387b) and argue that it works just as well in both cases. Consider the LF (388a) for our sentence (387a), which yields the truth conditions φ in (388b), which correspond to the unattested existential reading of the BPS 'firemen' that I want to rule out. Of course, (388) is the same as (330) considered above for (387b), only with a modal on top.

- (388) a. [must [[GEN_{*t*} C(*t*)] [firemen(*x*) tall(*t*)]]]
 b. $\varphi = \lambda w . \Box_{w' \in MB(w)} \text{GEN}_t [C^{w'}(t)] [\exists_x [[\text{fireman}]^{w'}(x) \wedge [\text{tall}]^{w'}(x, t)]]]$.

Consider next the LF (389a), obtained by replacing the BPS 'firemen' of LF (388a) with the definite description 'the fireman such and such', abbreviated as usual as 'the fireman P '. The corresponding truth conditions are ψ_P in (389b), where I am using the usual shorthand $d_P^{w'}$ for the individual [the fireman P]^{*w'*}. Since the BPS 'firemen' and the definite description 'the fireman such and such' are Horn-mates by assumption (333), then I can conclude that (389) is a scalar alternative of (388), according to the standard definition (74) of the set of scalar alternatives.⁴²

- (389) a. [must [[GEN_{*t*} C(*t*)] [the fireman P tall(*t*)]]]
 b. $\psi_P = \lambda w . \Box_{w' \in MB(w)} \text{GEN}_t [C^{w'}(t) \wedge \text{in}^{w'}(d_P^{w'}, t)] [[\text{tall}]^{w'}(d_P^{w'}, t)]]$.

Consider ψ as in (390), where \exists_x is a shorthand for the existential quantifier over firemen. Note that ψ only differs from φ in (388b) because of the fact that the existential quantifier over firemen \exists_x has wide scope over the generic operator GEN_{*t*}. Note the crucial difference between ψ in (390), that we get for the case of embedding under a universal modal, and ψ in (383), that we got above for Fox's case of embedding under a universal operator over individuals: in the latter case (383), the existential quantifier \exists_x had wide scope over the universal quantifier \forall_y (as well as over the generic operator); but in the former case (390), the existential quantifier \exists_x does not have wide scope over the universal modal (but only over the generic operator).

- (390) $\psi = \lambda w . \Box_{w' \in MB(w)} [\exists_x [\text{GEN}_t [C^{w'}(t) \wedge \text{in}^{w'}(x, t)] [[\text{tall}]^{w'}(x, t)]]]]$.

The logical equivalence in (391) holds,⁴³ in words: the negation of all the alternatives ψ_P 's in (389) is equivalent to the negation of ψ in (390), namely to the statement that "it is possible that there is no fireman who is tall throughout his entire life-span."

⁴²Furthermore, since ψ_P is not logically weaker than (but actually logically independent of) φ , then I can conclude that (389) is a scalar alternative of (388) also according to Fox's alternative definition (124) of the set of scalar alternatives.

⁴³First, let me show that if $\neg\psi$ is true in a world w , then $\bigwedge_P \neg\psi_P$ is true too in w . In fact, suppose by contradiction that the claim is false, namely that $\bigwedge_P \neg\psi_P$ is false in w ; this means that there is at least one property \tilde{P} such that $\psi_{\tilde{P}}$ is true in w ; this means in turn that, in every world $w' \in MB(w)$, the fireman such that \tilde{P} (namely the individual $d_{\tilde{P}}^{w'}$) is tall throughout his entire life span in w' ; thus, ψ is obviously true too in w , contradicting the initial hypothesis. Vice versa, let me show that if $\bigwedge_P \neg\psi_P$ is true in a world w , then $\neg\psi$ is true too in w . In fact, suppose by contradiction that the claim is false, namely that ψ is true in w ; this means that for each world $w' \in MB(w)$, there is a fireman $d = d^{w'}$ (possibly, a different one in different worlds) such that $d^{w'}$ is tall throughout his entire life-span; consider the property \tilde{P} such that, for every world $w' \in MB(w)$, 'the fireman such that \tilde{P} ' denotes this special fireman $d^{w'}$ (namely $d_{\tilde{P}}^{w'} = d^{w'}$); note that the truth of ψ in w thus entails the truth of the the corresponding $\psi_{\tilde{P}}$ in w and thus the falsity of $\bigwedge_P \neg\psi_P$ in w , which contradicts the initial hypothesis.

$$(391) \quad \bigwedge_P \neg\psi_P = \neg\psi$$

Let me now argue that the blind strengthened meaning $\text{EXH}(\varphi)$ of φ in (388) is the one in (392), i.e. that all the alternatives ψ_P 's in (389) are excludable w.r.t. φ . Given the equivalence in (391), it suffices to show that $\varphi \wedge \neg\psi$ is not a logical contradiction. And of course it is not, since it would be true for instance in a world w such that all the worlds w' accessible from w are as depicted in (337).

$$(392) \quad \text{EXH}(\varphi) = \varphi \wedge \underbrace{\bigwedge_P \neg\psi_P}_{\neg\psi}$$

Let me now state the technical assumption (393), namely that only worlds compatible with common knowledge are accessible from a world compatible with common knowledge.

$$(393) \quad \text{If } w \in \mathcal{W}_{\text{ck}} \text{ and } w' \in \text{MB}(w), \text{ then } w' \in \mathcal{W}_{\text{ck}}.$$

Under assumption (393), the strengthened meaning $\text{EXH}(\varphi) = \varphi \wedge \neg\psi$ that I just computed is a contradiction given common knowledge (227), as stated in (394); let me explain why. Consider a world $w \in \mathcal{W}_{\text{ck}}$ where φ is true. Hence, for each world w' accessible from w , there is at least a fireman who is tall at at least one instant in w' . Since w is compatible with common knowledge, then each such world w' is compatible with common knowledge too, by assumption (393). Thus, in each world w' accessible from w there is a fireman who is tall throughout his entire life span, i.e. ψ is true in w .

$$(394) \quad \text{EXH}(\varphi) \wedge \mathcal{W}_{\text{ck}} = \varphi \wedge \neg\psi \wedge \mathcal{W}_{\text{ck}} = \emptyset$$

In conclusion, the lack of the existential reading of the BPS of the modalized sentence (387a) is predicted in exactly the same way as the lack of the existential reading of the BPS of the plain sentence (387b).

3.6 More facts on the distribution of existential BPSs

At the beginning of this chapter, I have characterized the distinction between s- and i-predicates in terms of homogeneity with respect to the time argument. And in section 3.3, I have derived the generalization (395) from this characterization of the s-/i-predicates distinction. Let me refer to (395) as the STANDARD BPS-GENERALIZATION.

- (395) a. BPSs of i-predicates admit the generic reading but not the existential one;
 b. BPSs of s-predicates admit both the existential and the generic reading.

Yet, both claims of the standard BPS-generalization (395) have been challenged by various scholars. If these scholars are right, then I have an account for the wrong empirical generalization. In this section, I'll review various counterexamples to (395) and try to evaluate them. To anticipate, I will try to establish the following conclusions: that most of the counterexamples against claim (395a) reported in the literature do *not* really make the point they are intended to make; that there is a subset of these counterexamples that do make the point yet, and that I hope share the common property of being (physical or abstract) locatives in the sense of Gruber (1965); that most counterexamples to claim (395b) do make the point; that, nonetheless, these counterexamples to (395b) do not really argue against the BPS-generalization being stated in terms of homogeneity but rather against homogeneity being defined only in terms of the temporal argument.

3.6.1 First set of facts: existential BPSs that are non-kind denoting

Glasbey (1997, p. 169-170) considers the minimal pairs (396)-(398) and reports the intuition that the existential reading for the BPS is "very difficult, if not impossible" for the (a) sentences but "readily available" for the (b) sentences. Glasbey further notes that "the phenomenon [whereby the BPSs of the i-predicates in the (b) sentences admit the existential reading] is not a result of the predicates in

question not being truly *permanent* or i-level. She thus concludes that the (b) sentences provide a counterexample to the generalization (395a) that BPSs of i-predicates cannot receive the existential readings.

- (396) a. Students own sports cars.
 b. Students own sports cars in this department.
- (397) a. Drinkers were under-age.
 b. I was shocked to discover in the Red Lion last night that drinkers were under-age.
- (398) a. Ministers are gay.
 b. In this church, ministers are gay.

Some more examples of BPSs with alleged existential reading that have been offered in the literature as counterexamples to claim (395a) are reported in (399), from Cohen and Erteschik-Shir (2002) and Kiss (1998).

- (399) a. Chapters of this book are interesting.
 b. Family members are proud of John.
 c. In this area, hot springs exist.
 d. Fish abound in this lake.
 e. Men are bald as a result of using this hair restorer.

Finally, Fernald (2000, p. 98) notes that the BPSs in the (a) sentences in in (400)-(402) can only be construed generically while the one in the (b) sentences can be construed existentially: “there are a number of cases in which definite nominals seem to cause what we would otherwise interpret as an i-predicate to act like an s-predicate. [Consider for instance the pair in (400):] ‘own banks’ has all the characteristics of an i-predicate. On the other hand ‘own that house’ looks like an s-predicate. We must conclude that the nominal object of ‘own’ plays a crucial role in determining possible readings for indefinite subjects.” The (b) sentences thus provide a counterexample to the generalization (395a) that BPSs of i-predicates cannot receive the existential readings.

- (400) a. Tycoons own banks.
 b. Tycoons own that house.
- (401) a. Presidents are similar to senators.
 b. Presidents are similar to these senators.
- (402) a. Monkeys live in trees.
 b. Monkeys live in that tree.

Let me discuss these counterexamples to the standard BPS-generalization (395a) in some detail. I think that the examples in (399) are not interesting. The first two examples (399a) and (399b) are actually judged as deviant by my informants (Ezra Keshet and Tamina Stephenson (p.c.)). Kimball (1973, p. 268) notes that “for Frege, existence was a predicate, but of a special sort. It was a second order predicate, a predicate of predicates, not a predicate of objects. To say ‘Tame tigers exist’ for Frege was to say of the *concept* ‘tame tigers’ that something fell under it. Thus, the logical subject of such an utterance was not any tame tiger, but the concept itself.” After Carlson (1977), this passage can be reinterpreted saying that ‘exists’ is a kind-level predicate. McNally (1998b) shows how to implement this idea in an explicit semantics for ‘there’-sentences. If indeed ‘exist’ is a kind-level predicate, then example (399c) can be put aside. I would like to suggest that also the predicate ‘abound’ in sentence (399d) is a kind-level predicate, as shown by Carlson’s (1977; pp. 47-48) test that it resists having quantified NP’s as subjects.

- (403) a. #All / most / three / some / many / ... fish abound in the lake.

Finally, I would like to suggest that the availability of the existential reading for the BPS in (399e) is due to the fact that the sentence contains a hidden ‘became’ as is (404a). Under this analysis, (399e) contains the s-predicate ‘to become bald’ and can thus be put aside. Indeed, Asaf Bachrach

(p.c.) points out that the existential reading of the BPS ‘men’ in (399e) disappears if ‘as a result of’ is replaced by ‘despite’, as in (404b).

- (404) a. Men *became* bald as a result of using this brand of hair restorers.
 b. Men are bald despite using this brand of hair restorers.

Let me now turn to Fernald’s contrasts (400)-(402). The last example (402) seems to me dubious, since it is not clear to me why ‘to live’ should be considered an i-predicate. In any event, note that the contrast does not seem to me to extend to other i-predicates, as shown by the lack of contrast in the interpretation of the BPS in the analogous minimal pairs (405)-(407). I thus conclude that Fernald’s observation only holds for a small special class of i-predicates. This is of course the same conclusion reached in subsection 3.3.1 in the discussion of the JCS’s proposal. I have no account to offer as to what makes these i-predicates special in this respect.⁴⁴

- (405) a. Italians know good wine.
 b. Italians know this good wine.
 (406) a. Italians love good food.
 b. Italians love Barilla.
 (407) a. Firemen are taller than policemen
 b. Firemen are taller than Michel Jordan.

Let me finally turn to the minimal pairs in (396)-(398), and argue that the availability of the existential reading for the BPSs of the (b) sentences is actually compatible with the proposal I have presented so far.⁴⁵ Carlson (1977, § 5.4) points out the very interesting fact that not all (English) BPs

⁴⁴As pointed out to me by Kai von Fintel (p.c.), there seem to be an interesting parallelism between Fernald’s contrast (400) and the following contrast, noted in von Fintel (2003, p. 287): given the piece of common knowledge that French is a republic, sentence (ia) sounds like an instance of presupposition failure; but sentence (ib) does not, rather it sounds as a fine false sentence. As von Fintel notes, “the predicate here is an i-predicate in both cases. The difference lies in whether there is a specific object that the king of France is claimed to own.”

- (i) a. The king of France owns a pen. *(presupposition failure)*
 b. The king of France owns this pen. *(false, and no presupposition failure)*

von Fintel only discusses the i-predicate ‘own’. It is not clear to me whether his contrast extends to other transitive i-predicates.

⁴⁵In the discussion of these examples (396)-(398), I will assume that Glasbey and Cohen and Erteschik-Shir are on the right track in claiming that the BPSs in the (b) sentences have an existential reading. Yet, I am not completely sure that this claim is really correct. Cohen and Erteschik-Shir themselves note that what they call an existential reading doesn’t quite feel as a plain existential reading: “One may argue that [these] sentences imply stronger than simple existential claims, e.g. [(399c)] implies that more than a handful of family members are proud of John. We do, in fact, agree with this intuition, but in our opinion this is an implicature, rather than part of the meaning of the sentences [...]. Hence, this fact is not in contradiction with the claim that the BPs in these sentences are read existentially.” They refer to Cohen (2000) for discussion of this implicature; Cohen (2005) might be a later version of that proposal, but I actually don’t quite see how it might help in deriving the desired implicature. As a matter of fact, I think that what these authors really have in mind for many of these examples is something like the syllogism in (i).

- (i) a. These BPSs do not have the generic reading;
 b. a BPS can have either the generic or the existential reading;
 c. hence, these BPSs have the existential reading.

This syllogism is sometimes made explicit, as in the following passage from Cohen and Erteschik-Shir (2002, p.129-130): “What is the interpretation of the BPSs in these sentences? It is clearly not generic [...]. The reading of the[se] BPSs is, therefore, the existential reading.” I accept hypothesis (ia), that the BPSs in the (b) sentences in (396)-(398) do not have a true generic reading, as in ‘Firemen are tall’. But I deny hypothesis (ib), and thus the conclusion (ic) altogether. As a matter of fact, it has been pointed out by Condoravdi (1997) that, besides the generic and the existential reading, BPSs (of both i- and s-predicates) admit also a third reading, that she dubs the FUNCTIONAL READING. Condoravdi illustrates this third reading by means of examples such as (ii). She points out that, “intuitively, [(ii)] appears synonymous with [(iia)]”. Thus, its BPS has neither the existential nor the generic reading, i.e. sentence (ii) cannot be paraphrased neither with (iib) nor with (iic). Infact, “unlike [(iib)], [(ii)] does not make an existential assertion but, like [(iia)], it is an assertion about the totality of the contextually relevant students, whose existence in the actual world seems to be presupposed by both [(ii)] and [(iia)].” Furthermore, sentence (ii) crucially differs from (iic): “although the bare plural receives a universal reading, [(iia)] is not generic in any obvious way; it does not express a non-accidental generalization about students in general, nor a regularity

are alike. He writes: “we have so far been concentrating our attention on NP’s that make reference to kinds of things. However, there appears to be a rather enigmatic class of bare plural NP’s which do not seem to denote kinds as we have imagined them. This class of bare plural NP’s, so far as I know, always is modified by a relative clause or a PP.” A few examples are given in (408).

- (408) a. parts of that machine.
 b. people in the next room.
 c. books that John lost yesterday.
 d. bears that are eating now.

Carlson points out that non-kind-denoting BPs can be easily pulled apart from kind-denoting BPs by means of the following two tests. *First*, kind-level predicates cannot be predicated of non-kind-denoting BPs, as illustrated in (409a). *Second*, the two types of BPs differ with respect to their scope possibilities. As recalled in section 3.3, kind BPs always take the narrowest possible scope; but non-kind BPs can take wide scope. The classical contrasts (410)-(411) illustrate this difference: the existential BP in (410a) can only take narrow scope w.r.t. the modal, while sentence (410b) admits both an opaque and a transparent reading; analogously, the existential BP in (411a) can only have narrow scope w.r.t. negation, while sentence (411b) displays scope ambiguity.

- (409) a. ?People in the next room are common.
 b. ?People in the next room are numerous.
 c. ?People in the next room are indigenous to Asia.
 d. ?People in the next room are widespread.
- (410) a. John is looking for machines. $*\exists > \text{looking}, \text{looking} > \exists$
 b. John is looking for parts of that machine. $\exists > \text{looking}, \text{looking} > \exists$
- (411) a. John didn’t see machines. $*\exists > \neg, \neg > \exists$
 b. John didn’t see parts of that machine. $\exists > \neg, \neg > \exists$

I would like to suggest that the BPSs in (396a), (397a) and (398a) are kind BPs while those in (396b), (397b) and (398b) are non-kind BPs. As a matter of fact, the difference between Glasbey’s (a) and (b) sentences is just that a PP has been added, which can plausibly be construed as modifying the BPS (e.g. ‘students in this department’, ‘drinkers in the Red Lion’, etc.). Thus construed, these BPSs

about the occurrence of awareness in other situations in which a ghost was haunting the campus.”

- (ii) In 1985 there was a ghost haunting the campus. Students were aware of this fact / the danger.
 a. The students were aware of this fact / the danger.
 b. There were students who were aware of this fact / the danger.
 c. Students are usually aware of this fact / the danger.

Let me thus assume that the following test can be used to single out whether a BPS has the functional reading: replacement of the BPS with the corresponding definite description should not in any way affect the meaning of the sentence. It seems to me that the following (a) sentences, repeated from above, can be faithfully paraphrased by the (b) or (c) sentences, with the corresponding definite description.

- (iii) a. In this church, ministers are gay.
 b. The ministers of this church are gay.
 c. In this church, the ministers are gay.
- (iv) a. Students own sports cars in this department.
 b. The students of this department own sport cars.
 c. The students own sports cars in this department.
- (v) a. Family members are proud of John.
 b. His family members are proud of him.

I conclude that the BPSs of these (a) sentences do not have the existential reading but rather Condoravdi’s functional reading. Thus, these (a) sentences are not counterexamples to claim (395a) of the standard BPS-generalization, that BPSs of i-predicates lack the existential reading. I leave it open for the time being whether Condoravdi’s functional reading of BPSs is problematic for the proposal I am defending in this work.

cannot take kind-level predicates, as shown in (412); furthermore, they display scope freedom w.r.t. negation, as shown in (413).

(412) ?Students of this department are common/widespread.

- (413) a. John didn't see students. $*\exists > \neg, \neg > \exists$
 b. John didn't see students of this department. $\exists > \neg, \neg > \exists$

The account for the lack of the existential reading of BPSs of i-predicates that I have presented in section 3.3 crucially relied on the fact that BPSs can only take narrow scope w.r.t. the generic operator which binds the time argument of the i-predicate. Thus, my proposal only applies to kind BPSs but does not apply to non-kind BPSs, because of their free scope possibilities. In conclusion, my proposal predicts the contrasts in (396)-(398), given that the BPSs in the (b) sentences are non-kind BPs and thus nothing rules out their existential reading.

3.6.2 Second set of facts: BPSs associated with 'only'

von Stechow (1997) notes that the BPS of an i-predicate does get the existential reading when the BPS is associated with 'only': "it is obvious that speakers who utter these sentences [(414)] do not have to presuppose that professors in general are confident, [...], that all intelligent people are physicists [...]. So, if what we feel these sentences as signaling about the speaker's presupposition is a straightforward clue as to what the prejacent proposition is, we have to conclude that we are dealing with existentially quantified prejacent. And that would be so even though the putative prejacent uttered on their own are not readily understood existentially." Kiss (1998, p. 148) makes the same point: she considers sentence (415) and notes that it "can also be true about a village in which only a minority of women, but none of the men, have blue eyes." Thus, association with 'only' seems to be a suitable device to build counterexamples to claim (395a) of the standard BPS-generalization.⁴⁶

- (414) a. Only professors_F are confident.
 b. Only [intelligent people]_F are physicists.
 (415) In this village, only women_F have blue eyes.

Let me argue that these cases are fully compatible with my proposal. As recalled above, (kind) BPs always take the narrowest possible scope. But Fox et al. (2001) note that BPs associated with 'only' do take wide scope. They support this claim with the example (416): sentence (416a) sounds odd because the existential BP 'books' can only take narrow scope w.r.t. negation, and thus the sentence ends up saying that the censor has banned every book; in (416b) the same BP is associated with 'only' and the sentence feels fine, showing that the existential BP can scope above negation in this case. I have no proposal to make as to why BPs behave in this special way when they associate with 'only'.

- (416) We live in a regimented society where reading material is censored. The way this is done is that once a week this guy comes over and provides us with a list of books and articles that we are not allowed to read. The list is never the entire set of books or articles. In fact, there is always plenty of reading material in accord with the values of the society. One week there are no articles on the list.
 a. ?He said that we shouldn't read books.
 b. He only said that we shouldn't read BOOKS.

The proposal presented in section 3.3 only rules out the existential reading of (kind) BPSs of i-predicates which are condemned to narrow scope w.r.t. the generic operator which binds the time argument of the i-predicate. If the BPSs in (414) can have wide scope because of association with

⁴⁶Kiss' example (415) might partially be confounded by the locative 'in this village', which might make the BP non-kind denoting, as discussed in the preceding subsection (or else could make it receive Condoravdi's functional reading, as discussed in footnote 45).

'only' (for whatever reason), then the availability of their existential reading is fully compatible with my proposal.⁴⁷

3.6.3 Third set of facts: generalized locatives?

Kiss (1998) and McNally (1998a) provide the examples in (417) with locatives; see also Dowty (1979; p. 191, fn. 15) and Carlson (1977; par. 7.4). McNally comments that in these examples "the locative property is temporally persistent and yet the BPs are construable as weak existentials". In other words, BPSs of locative i-predicates admit the existential reading.

- (417) a. Volcanoes line both sides of the river.
 b. Large forests were on either side of the canyon.
 c. Cities lie along that river.
 d. Ruins were at the foot of the mountain.

Kratzer (1995) notes that the existential reading is available for the BPSs in (418). Kratzer concludes from (418) that BPSs of passive and unaccusative i-predicates admit the existential reading.

- (418) a. Counterexamples to this claim are known to me.
 b. Ponds belong to this property.

Yet, the BPSs of other passive i-predicates, such as those in (419a), do not admit the existential reading, contrary to (418a). Furthermore, there might just be something very special going on with the predicate 'belong' in (418b), independently of it being unaccusative: 'belong' is related to the i-predicate 'own', whose BPS surprisingly admits the existential reading, as in Fernald's example (400), repeated in (419b).

- (419) a. Italians are loved/hated by John.
 b. Italians own this restaurant.

As pointed out by Diesing (1992, p. 46), it is exciting to note that the recalcitrant cases of locatives in (417), 'know' in (418a) 'belong' in (418b) and 'own' in (419b) are exactly the cases which exemplify the generalized notion of locatives explored in Gruber (1965), as can be seen from the following quote from Jackendoff (1972): "with verbs of location, the Theme is defined as the NP whose location is being asserted. In (420a)-(420c), we are dealing with physical location, and 'the rock', 'John' and 'the book' are Themes. (420d)-(420f) involve possessional location, and 'the book' is theme in each case. (420g) is an abstract analogue of possession, so 'the answer' is Theme."

- (420) a. The rock stood in the corner.
 b. John clung to the window sill.
 c. Herman kept the book on the shelf.
 d. Herman kept the book.
 e. The book belongs to Herman.
 f. Max owns the book.
 g. Max knows the answer.

I thus tentatively suggest that these examples show that there is something very special going on in the case of BPSs of locative and extended locative i-predicates. I have no concrete proposal to make at this regard.⁴⁸

⁴⁷Kiss (1998) furthermore considers examples like (i) and observes that the BPS 'girls' can be construed existentially, i.e. the sentence feels true if no boy knows math and only *some* girls do.

(i) GIRLS know mathematics the best in my school.

He thus concludes that contrastive focus, more generally, licenses existential BPSs of i-predicates. To get this parallelism between association with 'only' and plain focus, it would be nice if the contrast in (416) still hold also when 'only' is replaced by contrastive focus. I am not clear about the judgments.

⁴⁸The case of locatives might indeed be very special. For instance, locatives provide a sharp exception to at least another well known generalization concerning the i- vs s-predicates distinction. Portuguese has two copulas, 'ser' and 'estar'.

3.6.4 Fourth set of facts: statives

Also the second half (395b) of the Standard BPS-Generalization has been criticized in the literature. Evidence against claim (395b) of the standard BPS-generalization comes from the fact that *i*-predicates coerced into an *s*-level meaning still do not allow their BPSs to have an existential reading. This point has been made for instance in Condoravdi (1992), McNally (1998a) and Jäger (2001). For instance, McNally considers the examples (421) and (422) and comments as follows: in the context described in (a), “the (b) examples below can have a reading that corresponds to quantification over eventualities, a reading that only *s*-predicates are generally assumed to license [...]. In such cases, Kratzer (1995) and others have suggested that *i*-predicates are COERCED into behaving as stage level counterparts. Nonetheless, the bare plural in the (c) sentences cannot easily be construed as weakly existential, as one might expect under coercion. For example, it is not sufficient for the truth of (422c) that only two or three of the daffodils in the context are suddenly tall; instead, the subordinate clause most naturally expresses a generalization about the contextually relevant set of daffodils at some spatiotemporal index. [...] The failure of coercion to affect the interpretation of the bare plurals in the (c) examples indicates that the temporal properties of a state or the way in which it is individuated in time do not affect the interpretation of the bare plural”.

- (421) a. In this town, firemen dye their hair every other day.
 b. (When) blond, they have more fun.
 c. Today, firemen are blond. * \exists -BPS
- (422) a. The characters in “Alice in Wonderland” are constantly changing size at will by eating and drinking various things.
 b. (When) tall, Alice can’t get through that tiny wooden door.
 c. Alice knew that her animal friends could change size, but she was surprised to learn that daffodils were suddenly very tall. * \exists -BPS

The main piece of evidence against claim (395b) of the standard BPS-generalization has to do with adjectival predicates. For example, Kiss (1998) writes: “There are *s*-predicates that do not allow [...] existentially interpreted BPSs. Such are almost all adjectival predicates. [...] Even though these predicates express temporary, stage-level properties, their subjects cannot be interpreted existentially.” Analogously, Higginbotham and Ramchand (1997) write that “generic interpretation is the norm for predication of bare plurals in the simple present, rather than a feature that attaches to *i*-predicates” since “if we survey examples over than the well known [...], we find that [BPSs of] predicates of a generally transient nature resist existential interpretation.” Some of these examples are quoted in (423) and (424), from Higginbotham and Ramchand (1997), Cohen and Erteschik-Shir (2002), Kiss (1998) and McNally (1998a). McNally comments on the examples in (424) as follows: despite the fact that the predicates in these examples “express transient states”, “it is difficult even with the facilitating context to interpret the bare plurals in these sentences as weak existentials. Rather these sentences [...] express generalizations over the denotations of their subjects in some temporally constrained domain (if indeed they sound felicitous at all).”

- (423) a. Children are fat / skinny / sick / joyful / sad.

Roughly, ‘*estar*’ patterns with *s*-predicates (it cannot be used with true *i*-predicates, such as ‘has blue eyes’; furthermore, it yields a temporary reading of neutral predicates, such as ‘happy’); ‘*es*’ patterns with *i*-predicates (it cannot be used with true *s*-predicates, such as ‘is pregnant’; furthermore, it yields a permanent reading for neutral predicates). As Schmitt (2005) points out, locatives provide a blatant counterexample to this generalization: sentences (i) express permanent properties, and nonetheless ‘*estar*’ is perfectly acceptable.

- (i) a. A casa está no fim da rua.
 The house ESTAR in-the end of-the street.
 ‘The house is in the end of the street’
 b. As Montanhas Rochosas estão no Colorado.
 The Rocky Mountains ESTAR in-the Colorado
 ‘The Rocky Mountains are in Colorado.’

- b. Stars are old / young.
 - c. People are rich.
 - d. Shoes are shiny.
 - e. Boys are brave.
 - f. People are on holiday.
 - g. People are at work.
 - h. People are on coffee break.
 - i. Girls are good at learning foreign languages.
- (424)
- a. Committee members were bored until the Dean suddenly showed up.
 - b. Today, people in the office were in a good mood.
 - c. During the class, farmers were hungry / tired / cheerful.
 - d. Yesterday butter was old / fresh.
 - e. The diners complained because plates were dirty / greasy.
 - f. Turn on the dryer again because shirts are still damp.

Kiss (1998) adds the following contrast: contrary to the adjectival predicate ‘to be noisy’ in (425a), “the verbal predicate corresponding to ‘noisy’ allows the existential reading of its BPS”, as shown in (425b). Overall, these examples show that lack of the existential reading of BPSs has nothing to do with the temporal properties of the predicate, contrary to what asserted by claim (395b) of the standard BPS-generalization.⁴⁹

- (425)
- a. Children are noisy. * \exists -BPS
 - b. Children are making noise. \exists -BPS

In light of the examples (423) and (424), the examples in (421) and (422) don’t really say much: coercion transforms the i-predicates ‘blond’ and ‘tall’ into stative adjectives which behave just as those in (423) and (424) with respect to the readings of their BPSs. The crucial counterexamples against claim (395b) of the standard BPS-generalization are thus those in (423) and (424): non persistent stative adjectives whose BPSs lack the existential reading. What’s going on here? Here is a very tentative suggestion. Following Chierchia (1995), I have suggested that i-predicates are inherently generic, namely that they have a davidsonian argument which always ends up bound by a generic operator, as in the truth conditions (276b) repeated in (426), analogous to those of habitual s-predicates. To the extent that statives in general pattern like i-predicates, I thus have to maintain that statives too are inherently generic, namely that their davidsonian argument is always bound by a generic operator. I am not in a position to really defend this hypothesis on statives here. Yet, I would like to suggest that it might not be implausible.

- (426) a. John is tall.

⁴⁹There are two interesting contrasts discussed in Higginbotham and Ramchand (1997), that I have nothing to say about. One contrast is (i): the BPS of the matrix clause (ia) can be construed existentially but the BPS of the embedded clause (ib) does not seem to admit an existential construal.

- (i)
- a. Firemen are available / on strike.
 - b. I consider firemen available / on strike.

Another contrast is (ii): that the BPS of sentence (iia) admits the existential reading but the one of sentence (iib) does not.

- (ii)
- a. (Guess whether) firemen are nearby / at hand.
 - b. (Guess whether) firemen are far away / a mile up the road.

Higginbotham and Ramchand (1997) seem to make a big point out of the latter contrast (ii). Indeed, they write: “many of these predicates [which allow for an existential BPS] involve the spatio-temporal proximity of the subject to the speaker. Thus contrast (iia) with (iib). It seems that an insight into the issues here must take account of some notion involving the local scene” (p. 66); and also: “although it would take us too far afield to consider all of the data here, the predicates that naturally allow existential readings seem to us generally to involve, not merely spatio-temporal location, but spatio temporal proximity to the speaker. This special class does not appear to be structurally distinguished in any way” (p. 56).

- b. $\text{GEN}_t[\text{in}(j, t)][[\text{tall}]](j, t)$

Indeed, this hypothesis on statives is very close in spirit to the old characterization of statives in terms of the SUBINTERVAL PROPERTY suggested by Bennett and Partee (1972). This characterization says that, whenever a stative holds of an individual at a given interval, it holds of that individual at any subinterval. As a matter of fact, this characterization of statives sounds very similar to my characterization (227) of an *i*-predicate such as ‘tall’, according to which, if it holds of an individual at a given time, it always holds of that individual. Another way to make sense of the hypothesis that statives are inherently generics is to consider their behavior w.r.t. modification. *I*-predicates such as ‘tall’ do not allow *temporal* modification, as shown in (427a). The stative adjectives in (423) of course do, as shown in (427b). Yet, these stative adjectives do not allow *locative* modification, as shown in (427c); see Maienborn (2001) for arguments in favor of this latter claim.

- (427) a. ?Yesterday, John was tall.
 b. Yesterday, John was sleepy.
 c. ?John was sleepy in his car.

In Subsection 3.2, I have interpreted the oddness triggered by the tense modification in (427a) as evidence for assumption (227) that ‘tall’ is persistent w.r.t. time. Since (427b) is fine, then these stative adjectives are not persistent w.r.t. time. Throughout this paper, I have construed Davidsonian arguments in the simplest possible way, namely as times. With this choice, there is no way to get the analogy between the ban against temporal modification in (427a) and locative modification in (427c). Yet, this assumption on davidsonian arguments is of course an oversimplification. Thus, assume that davidsonian arguments have a richer ontological nature, say they are spatio-temporal situations. Now, I could get the analogy between (427a) and (427c) by assuming that these stative adjectives which do not allow locative modification are persistent too, only not w.r.t. time but rather, say, w.r.t. the space coordinate of their davidsonian argument. By analogy with (426), I could then posit something like (428). See for instance Deal (2008) for analogies between time and spatial locations.

- (428) a. John is sleepy.
 b. $\text{GEN}_\ell[\text{in}(j, \ell)][[\text{sleepy}]](j, \ell)$

I could thus repeat for the BPSs of the adjectives in (423) the account presented in section 3.3 for the BPSs of *i*-predicates, replacing the time argument with situations or locations. This line of reasoning of course predicts that, if a stative does allow for local modification, then its BPS is not prevented from getting the existential reading. And the data in (429) indeed show that those statives that do admit locative modification also allow their BPS to be construed existentially.

- (429) a. Firemen are available on Cambridge street.
 b. Stars are visible on the north portion of the sky.
 c. Students were present at the meeting.

Furthermore, this line of reasoning would predict a Fox’s effect analogous to (369) also for the BPSs of the adjectives in (423): embedding under a universal operator should rescue the existential reading of the BPS of these statives. As a matter of fact, the existential reading of the BPS ‘children’ seems more readily available in (430b) than in (430a).

- (430) a. Children are sick.
 b. Every day, children are sick.

I thus tentatively conclude that the examples in (423) do not really argue against correlating the lack of existential BPSs with persistence of the predicate but rather against persistence being construed only in terms of time. But I leave it open for the time being how exactly this intuition should be spelled out.

3.7 Overt universal Q-adverbs

As shown in (431), i-predicates cannot occur with an overt universal Q-adverb such as ‘always’ or ‘often’, contrary to s-predicates.

- (431) a. #John is always tall.
b. John is always available.

The oddness of sentence (431a) is particularly puzzling for my proposal. In fact, consider sentence (431a), repeated below in (432a), next to the fine sentence (432b) without overt ‘always’. The obvious LF of the odd sentence (432a) is φ in (433a). At the end of section 3.2, I have suggested that the LF of the fine sentence (432b) is ψ in (433b), with the generic operator GEN. Under the assumption that the quantificational force of ‘always’ and the generic operator GEN are the same, the contrast in (432) is surprising.

- (432) a. #John is always tall.
b. John is tall.
- (433) a. $\varphi = [\text{always}_t [\text{John tall}(t)]]$.
b. $\psi = [\text{GEN}_t [\text{John tall}(t)]]$.

In this section, I argue that the contrast in (432) can be accounted for in the same way in section 2.4 I have accounted for the contrasts in (171)-(175), one of which is repeated in (434).

- (434) In a context where all children inherit the last name of their father:
a. ?Every child of couple C has a French last name.
b. The children of couple C have a French last name.

My argument exploits the following parallelism: sentence (434a) competes with sentence (434b), where the universal quantifier is replaced by a definite; and this latter sentence (434) triggers a homogeneity presupposition; in the same way, sentence (432a) competes with sentence (432b), where ‘always’ is replaced by the covert generic operator GEN; and this latter sentence (432b) triggers a homogeneity presupposition too. For a discussion of the parallelism between the definite article and the generic operator, see for instance Ferreira (2005). To stress the analogy between the two accounts for the contrasts in (432) and (434), Table 3.8 summarizes the main steps of the two accounts, one next to the other.

Account Of course, the two LFs ψ and φ in (433) have the same meaning; nonetheless, they crucially have different presuppositions. As argued for example in von Stechow (1997), the covert generic operator GEN carries the *homogeneity presupposition* in (435), namely that the nuclear scope of GEN holds of each item in its restrictive clause or else is false for each such item. The homogeneity presupposition can be detected by means of negation: the sentence ‘It’s false that John smokes’ conveys that he never smokes, which is different from the plain meaning (namely that “it is false that John always smokes”) but does follow from the plain meaning plus the homogeneity presupposition. No such presupposition is carried by overt ‘always’.

- (435) $[[\text{GEN}_{v_1, \dots, v_n} [\alpha(v_1, \dots, v_n)] [\beta(v_1, \dots, v_n)]]] = \lambda w : \text{YES}^w \vee \text{NO}^w . \text{YES}^w$
a. $\text{YES}^w = 1$ iff $\beta^w(v_1, \dots, v_n)$ for all v_1, \dots, v_n s.t. $\alpha^w(v_1, \dots, v_n)$;
b. $\text{NO}^w = 1$ iff $\neg\beta^w(v_1, \dots, v_n)$ for all v_1, \dots, v_n s.t. $\alpha^w(v_1, \dots, v_n)$.

In the case of ψ in (433b), the generic operator binds a single variable, namely time; thus, the homogeneity presupposition of ψ according to (435) boils down to (436b), namely that “either John is tall at every time he is alive or he is never tall.” No such presupposition is carried by φ in (433a), as stated in (436a), where I am ignoring other potential presuppositions irrelevant to my point. Let me assume that ‘always’ and GEN are Horn-mates. Of course, ψ_{prs} asymmetrically entails φ_{prs} . Thus, the blind strengthened presupposition of φ boils down to $\text{EXH}_{\text{prs}}(\varphi) = \varphi_{\text{prs}} \wedge \neg\psi_{\text{prs}} = \neg\psi_{\text{prs}}$. This strengthened presupposition is a contradiction given common knowledge \mathcal{W}_{ck} in (227), since \mathcal{W}_{ck} entails ψ_{prs} , namely that John is either always tall or else he never is. The oddness of sentence (431a) is thus predicted.

?Every the children of that couple have a foreign last name.	?John is always tall.												
(1) LF AND PRESUPPOSITIONS													
<p> φ every children VP has a last name... $\varphi_{prs} = \mathcal{W}$ </p>	<p> φ always C John is tall $\varphi_{prs} = \mathcal{W}$ </p>												
(2) ASSUMPTION ON HORN SCALES													
<'every child', 'the children'>	<'always', GEN>												
(3) SET OF SCALAR ALTERNATIVES													
$Alt(\varphi_{prs}) = \{\psi_{prs}\}$, where ψ_{prs} is the presupposition of ψ defined as follows													
<p> ψ the children VP have a last name... </p>	<p> ψ GEN C John is tall </p>												
Both the definite and the generic operator bear the <i>homogeneity presupposition</i> ; see Fodor (1970), Löbner (1985), von Stechow (1997), Gajewski (2005), Ferreira (2005)													
$\psi_{prs} = \lambda w . [\forall x [[\text{VP}]^w(x)]] \vee [\forall x [\neg [[\text{VP}]^w(x)]]]$ $\psi_{prs} = \lambda w . [\text{GEN}_t [[\text{tall}]^w(j, t)]] \vee [\text{GEN}_t [\neg [[\text{tall}]^w(j, t)]]]$													
(4) SET OF EXCLUDABLE ALTERNATIVES													
$Excl(\varphi_{prs}) = \{\psi_{prs}\}$, since $\varphi_{prs} \wedge \neg \psi_{prs}$ is true e.g. in the following world													
<table border="1" style="border-collapse: collapse; text-align: center;"> <tr> <td></td> <td>John</td> <td>Bill</td> <td>Tom</td> </tr> <tr> <td>foreign:</td> <td>✓</td> <td></td> <td>✓</td> </tr> <tr> <td>not-foreign:</td> <td></td> <td>✓</td> <td></td> </tr> </table>		John	Bill	Tom	foreign:	✓		✓	not-foreign:		✓		$\text{C}^w(\cdot) \wedge \text{in}^w(j, \cdot)$ <hr style="width: 50%; margin: 0 auto;"/> $[[\text{tall}]^w(j, \cdot)]$
	John	Bill	Tom										
foreign:	✓		✓										
not-foreign:		✓											
(5) CONCLUSION													
The blind strengthened presupposition is $\text{EXH}(\varphi_{prs}) = \varphi_{prs} \wedge \neg \psi_{prs} = \neg \psi_{prs}$, which contradicts common knowledge													

Table 3.8: Parallelism between the account for the oddness of sentence (434a) presented in section 2.4 and the account for the oddness for sentence (431a) presented in this section.

- (436) a. $\varphi_{\text{prs}} = \mathcal{W}$.
 b. $\psi_{\text{prs}} = \text{YES} \cup \text{NO}$,
 i. $\text{YES} = \{w \mid \llbracket \text{tall} \rrbracket(j, t), \text{ for every } t \text{ such that } \text{in}^w(j, t)\}$,
 ii. $\text{NO} = \{w \mid \neg \llbracket \text{tall} \rrbracket(j, t), \text{ for every } t \text{ such that } \text{in}^w(j, t)\}$.

As already noted in section 3.1, Kratzer (1995) points out that sentence (431a), repeated once more in (437a), turns fine if the definite ‘John’ is replaced by either a BP or an indefinite, as in φ in (437b).

- (437) a. ?John is always tall.
 b. $\varphi = \text{Firemen / a fireman are / is always tall}$.

Let me show how the contrast in (437) follows from my proposal. Consider the alternative ψ in (438), where the overt Q-adverb ‘always’ has been replaced by the covert generic operator GEN.

- (438) $\psi = \text{Firemen / a fireman are / is GEN tall}$.

In the case of φ and ψ in (437b) and (438), the Q-adverbs GEN and ‘always’ are quantifying both over times and over firemen; see Lewis (1975), Heim (1988) and Diesing (1992) among others for a discussion of unselective binding. Thus, the homogeneity presupposition for ψ in (438) becomes (439b) in this case. Again, ψ_{prs} asymmetrically entails φ and the strengthened presupposition of φ boils down to $\text{EXH}_{\text{prs}}(\varphi) = \varphi_{\text{prs}} \wedge \neg \psi_{\text{prs}} = \neg \psi_{\text{prs}}$. But this time, $\neg \psi_{\text{prs}}$ is not a contradiction given the common knowledge \mathcal{W}_{ck} in (227), since there are of course worlds compatible with common knowledge where ψ_{prs} is false because some firemen are tall while others are not. The felicity of sentences (437b) is thus predicted.

- (439) a. $\varphi_{\text{prs}} = \mathcal{W}$.
 b. $\psi_{\text{prs}} = \text{YES} \cup \text{NO}$,
 i. $\text{YES} = \left\{ w \mid \begin{array}{l} \llbracket \text{tall} \rrbracket(x, t), \text{ for every } x, t \text{ s.t.} \\ \llbracket \text{firemen} \rrbracket^w(x) \text{ and } \text{in}^w(x, t) \end{array} \right\}$,
 ii. $\text{NO} = \left\{ w \mid \begin{array}{l} \neg \llbracket \text{tall} \rrbracket(x, t), \text{ for every } x, t \text{ s.t.} \\ \llbracket \text{firemen} \rrbracket^w(x) \text{ and } \text{in}^w(x, t) \end{array} \right\}$.

This account furthermore predicts that sentences of the type of (437b) — with a BPS, an i-predicate and ‘always’ — should sound odd in cases where common knowledge entails that the individuals denoted by the BPS are homogeneous with respect to the property denoted by the i-predicate. This prediction seems to be borne out by the contrast in (440), pointed out to me by Danny Fox (p.c.): sentence (440a) with overt ‘always’ sounds odd despite the presence of the BPS (if ‘a difficult language’ is given wide scope and taken to refer to Hebrew).

- (440) a. $\varphi = \text{?Hebrew speakers always know a very difficult language}$.
 b. $\psi = \text{Hebrew speakers know a very difficult language}$.

The presuppositions of φ and ψ in (440) are as in (441). Again, the strengthened presupposition of φ in (440a) entails $\neg \psi_{\text{prs}}$ in (441b), which in this case does mismatch with common knowledge, since common knowledge entails that all Hebrew speakers speak the same language, which is either hard (hence YES is true) or easy (hence NO is true).

- (441) a. $\varphi_{\text{prs}} = \mathcal{W}$.
 b. $\psi_{\text{prs}} = \text{YES} \cup \text{NO}$,
 i. $\text{YES} = \left\{ w \mid \begin{array}{l} \llbracket \text{know} \dots \rrbracket(x, t), \text{ for every } x, t \text{ s.t.} \\ \llbracket \text{Hebrew speaker} \rrbracket^w(x) \text{ and } \text{in}^w(x, t) \end{array} \right\}$,
 ii. $\text{NO} = \left\{ w \mid \begin{array}{l} \neg \llbracket \text{know} \dots \rrbracket(x, t), \text{ for every } x, t \text{ s.t.} \\ \llbracket \text{Hebrew speaker} \rrbracket^w(x) \text{ and } \text{in}^w(x, t) \end{array} \right\}$.

The case of sentence (440) just considered is of course analogous to the case of sentence (245c) considered in section 3.1.

Part II

Constraint promotion and
the OT on-line model for the
acquisition of phonology

Abstract of Part II — The second half of this dissertation develops a theory of update rules for the OT on-line algorithm that perform constraint promotion too, besides demotion: it motivates the need for such rules; it develops techniques to prove their finite time convergence; and it starts the investigation of their properties. *Chapter 4* explains why we need constraint promotion. More in detail, this chapter introduces the OT on-line model for the acquisition of phonology; it systematizes within a unified framework various existing results on update rules for the OT on-line model, underscoring the fact that no convergent update rules that perform promotion too have been proposed so far in the literature; finally, it motivates the need for update rules that perform promotion too by arguing that demotion-only update rules are unsuited to model the early stage of the acquisition of phonology prior to morphological awareness, when the learner can only posit fully faithful underlying forms. *Chapter 5* shows how to get constraint promotion. More in detail, this chapter presents two different techniques to prove finite time convergence for a variety of update rules that perform promotion too. One technique shares the combinatoric flavor of Tesar and Smolensky's analysis of demotion-only update rules. The other technique is very different, and consists of a strategy to adapt to OT results from the theory of on-line algorithms for linear classification. The latter technique has various consequences interesting on their own, explored in *Chapter 8*. In particular, it entails that linear OT has no computational advantage over standard OT, contrary to what has been recently suggested by various scholars. *Chapter 6* starts the investigation of the properties of update rules that perform promotion too, concentrating on the characterization of the final vector. More in detail, this chapter presents an invariant on the (dual) vector entertained by the OT on-line algorithm with a promotion-demotion update rule; it shows how to use this invariant in order to characterize the final vector entertained by the algorithm, by discussing a few examples; finally, it puts forward the conjecture that the strong internal symmetry of phonotactics comparative tableaux might enable the OT on-line model run with a promotion-demotion update rule to learn much of the phonotactics of the target language. *Chapter 7* carries further the investigation of promotion-demotion update rules, concentrating on the number of updates. More in detail, this chapter shows that the number of updates required by update rules that perform promotion too may be large and it attempts at making sense of this fact by showing that the problem of the acquisition of phonology, construed as the OT Subset problem, is NP-complete.

Chapter 4

Why we need constraint promotion

A number of computational models for the acquisition of phonology within OT have been put forward in the recent literature. In the second part of my dissertation, I concentrate on the OT ON-LINE MODEL, see for instance Tesar and Smolensky (1998, 2000), Boersma and Levelt (2000) and Curtin and Zuraw (2002). This is the simplest and most primitive among existing models. And it is also the most widely used (or, at least, assumed) model in the non-computational acquisitional literature. It therefore deserves a close investigation. In a nutshell, the idea of the model is as follows: the learner holds at every time a current ranking; gets a piece of data; consequently updates its current ranking by promoting and/or demoting certain constraints. Section 4.2 describes the general shape of the model. The crucial ingredient of the model is the update rule used to move from one ranking to the next. There has been an intense discussion on update rules for the on-line model in the recent computational OT literature. One of the hot issues of this literature is whether constraint re-ranking should happen by demotion only or by promotion only or by both promotion and demotion. Various successful demotion-only update rules have been presented in the literature, while the search for update rules that perform promotion too has failed so far. Sections 4.2.2-4.2.6 offer a detailed review of the state of the art on update rules for the OT on-line model. There has been less discussion of the issue of constraint promotion versus demotion in the empirical acquisitional literature. Fikkert and De Hoop (2009) raise the question: “Does the (re)ranking of constraints involve the demotion of markedness constraints, the promotion of faithfulness constraints, or can it be achieved by both the demotion and the promotion of constraints?” (p. 311). But they leave the issue open and seem to suggest that the issue only has theoretical significance: “Although in practice the two approaches [constraint promotion and constraint demotion] are highly similar, conceptually they are different” (p. 319). Gnanadesikan (2004) explicitly considers the hypothesis of constraint promotion: “The process of acquisition is one of promoting the faithfulness constraints to approximate more and more closely the adult grammar, and produce more and more marked forms. The path of acquisition will vary from child to child, as different children promote the various faithfulness constraints in different orders” (p. 73). But she does not provide arguments in favor of her hypothesis that learning happens through constraint promotion, nor does she make explicit the details of the promotion update rule she is positing. Bernhardt and Stemberger (1998) and Stemberger and Bernhardt (1999) too explicitly defend constraint promotion: “We are unsure as to how constraints are generally re-ranked. They may always be re-ranked higher. [...] We suggest that the typical way that children learn the ranking of constraints is to re-rank faithfulness constraints so that faithfulness increases” (1999). They defend their claim in favor of constraint promotion by discussing a few specific cases; see also Stemberger et al. (1999) and Stemberger and Bernhardt (2001). In section 4.3.4, I take up this issue and provide a very general and straightforward argument in favor of update rules that perform both promotion and demotion. The argument in a nutshell is that demotion-only is not suitable in order to model the early stage of the acquisition of phonology as described in Hayes (2004), namely the stage prior to morphological awareness, when the learner cannot take advantage of alternations and thus can plausibly only posit fully faithful underlying forms. In conclusion, this chapter opens the research project of devising and investigating update rules for the OT on-line model that perform

promotion too. The rest of this dissertation will lay out the beginning of this research project: chapter 5 will show how to get constraint promotion and chapters 6 and 7 will show how to study the corresponding OT on-line algorithm.

4.1 OT preliminaries

This section provides background definitions and results on the framework of standard OT, as defined in Prince and Smolensky (2004). Subsection 4.1.1 quickly reviews the basic framework, in order to establish the notation. Subsection 4.1.2 introduces a better representation for OT data in terms of comparative rows, after Prince (2002) and Tesar (1995). Finally, subsection 4.1.3 introduces a better representation for OT hypotheses in terms of ranking vectors, after Boersma (1997).

4.1.1 OT basics

I will denote by \mathcal{X} the set of UNDERLYING FORMS and by \mathcal{Y} the set of SURFACE FORMS. I will denote by $Gen : x \in \mathcal{X} \mapsto Gen(x) \subseteq \mathcal{Y}$ the GENERATING FUNCTION. Finally, I will denote by \mathcal{C} the CONSTRAINT SET. I will always assume that there is a finite¹ number n of constraints $C_1, \dots, C_n : (x, y) \in \mathcal{X} \times Gen(x) \mapsto \mathbb{N}$. These four ingredients $(\mathcal{X}, \mathcal{Y}, Gen, \mathcal{C})$ are called the UNIVERSAL SPECIFICATIONS of a typology. An example of universal specifications is provided in (1): the set of underlying forms \mathcal{X} and the set of surface forms \mathcal{Y} coincide; the generating function Gen is only allowed to modify voicing, but does not perform either deletion or epenthesis; the constraint set \mathcal{C} contains a markedness constraint against voiced obstruents and two variants of the faithfulness constraint for voicing, a general and a positional one.

$$(1) \quad \begin{array}{l} \text{a. } \mathcal{X} = \{ /ta/, /da/, /rat/, /rad/ \} \\ \text{b. } \mathcal{Y} = \{ [ta], [da], [rat], [rad] \} \\ \text{c. } Gen(/ta/) = Gen(/da/) = \{ [ta], [da] \} \\ \quad \quad \quad Gen(/rat/) = Gen(/rad/) = \{ [rat], [rad] \} \\ \text{d. } \mathcal{C} = \left\{ \begin{array}{l} F_{\text{pos}} = \text{IDENT[VOICE]/ONSET,} \\ F_{\text{gen}} = \text{IDENT[VOICE],} \\ M = * [+VOICE, -SONORANT] \end{array} \right\} \end{array}$$

Given an underlying form $x \in \mathcal{X}$, let $Gen(x)$ be the corresponding set of CANDIDATES. Consider two candidates $y^*, y \in Gen(x)$ and let's call y^* a WINNER candidate and y a LOSER candidate. The basic data units in OT are UNDERLYING/WINNER/LOSER FORM TRIPLETS (x, y^*, y) as in (2a): the first item of the triplet is the underlying form x , the second item is the intended winner candidate y^* , and the third item is a loser candidate y . A concrete example of such a triplet is provided in (2b) for the case of the typology in (1).

$$(2) \quad \begin{array}{cc} \text{a.} & \text{winner} \\ & | \\ & (x, y^*, y) \\ & | \\ & \text{loser} \\ \text{b.} & \text{winner} \\ & | \\ & (/rad/, [rat], [rad]) \\ & | \\ & \text{loser} \end{array}$$

An hypothesis in OT is a RANKING, namely a linear order \gg on the constraint set \mathcal{C} . An example of ranking over the constraint set in (1d) is provided in (3): this ranking sandwiches the markedness constraint in between the two faithfulness constraints, with the positional faithfulness constraint ranked at the top. Given an arbitrary ranking \gg , I can assume without loss of generality that it is

¹It turns out that this hypothesis that the constraint set be finite is in no way necessary. The theory developed in this work extends straightforwardly to a framework with an *infinite* number of constraints C_1, C_2, \dots , as long as any pair (x, y) of an underlying form $x \in \mathcal{X}$ and a corresponding candidate $y \in Gen(x)$ can violate only a *finite* number n of constraints, namely $C_k(x, y) = 0$ for all constraints C_k but at most a finite number of them.

$C_1 \gg C_2 \gg \dots \gg C_n$, since the numerical labels assigned to the constraints are arbitrary to start with.

$$(3) F_{\text{pos}} \gg M \gg F_{\text{gen}}$$

The definition of OT-COMPATIBILITY in (4a) links together a data unit, namely an underlying/winner/loser form triplet, with an hypothesis, namely a ranking. For instance, the ranking in (3) is OT-compatible with the triplet in (2b) according to the definition (4a).²

- (4) a. A ranking \gg is called OT-COMPATIBLE with an underlying/winner/loser form triplet (x, y^*, y) iff the highest \gg -ranked constraint that distinguishes between the two pairs (x, y^*) and (x, y) assigns fewer violations to the pair (x, y^*) than to the pair (x, y) .
- b. A ranking \gg is called OT-COMPATIBLE with a set of underlying/winner/loser form triplets iff it is OT-compatible with each of them according to (4a).
- c. A set set of underlying/winner/loser form triplets is called OT-COMPATIBLE iff it is OT-compatible with at least one ranking according to (4b).

The notion of OT-compatibility in (4) has the following important property: given a ranking \gg and the set of candidates $Gen(x)$ corresponding to an underlying form x , there exists one and only one candidate $y^* \in Gen(x)$ such that \gg is OT-compatible with the underlying/winner/loser form triplet (x, y^*, y) for every candidate $y \in Gen(x)$ different from this designated candidate y^* . Thus, the OT-GRAMMAR $OT_{\gg} : x \in \mathcal{X} \mapsto Gen(x)$ corresponding to a ranking \gg can be defined as in (5) for any underlying form x . Note that all OT-grammars corresponding to the same universal specifications are assumed to have the same domain \mathcal{X} , according to the RICHNESS OF THE BASE assumption.

$$(5) OT_{\gg}(x) = y^* \text{ iff } \gg \text{ is OT-compatible with the underlying/winner/loser form triplet } (x, y^*, y) \text{ for every candidate } y \in Gen(x) \text{ different from } y^*.$$

For instance, the OT-grammar corresponding to the ranking in (3) is described in (6). Since F_{pos} is \gg -ranked above M , it lets /da/ surface faithfully. Since M is ranked above F_{gen} , it neutralizes the final voicing of /rad/.

$$(6) \quad \begin{array}{ll} OT_{\gg}(/ta/) = [ta] & OT_{\gg}(/da/) = [da] \\ OT_{\gg}(/rat/) = [rat] & OT_{\gg}(/rad/) = [rat] \end{array}$$

The LANGUAGE $\mathcal{R}(OT_{\gg})$ generated by the ranking \gg is the range of the function OT_{\gg} , namely the set of those surface forms $y \in \mathcal{Y}$ that are attainable through \gg , namely such that there exists at least one underlying form $x \in \mathcal{X}$ such that the OT-grammar OT_{\gg} maps that underlying form x into that surface form y . The language corresponding to the ranking \gg in (3) is the one in (7).

$$(7) \{[da], [ta], [rat]\}$$

Before I conclude, let me note that there are of course alternative, more direct ways of defining an OT grammar OT_{\gg} , besides (5). For instance, one might define $OT_{\gg}(x)$ as the maximum over $Gen(x)$ w.r.t. a properly defined total order induced by the ranking \gg . Here, I have chosen a more indirect way, whereby the notion of OT-compatibility (4) takes the foreground, and the notion of OT grammar is derived from that, as in (5). As it will become clear as we proceed, this way is best suited to the perspective adopted in this work.

²The notion of OT compatibility (4a) can also be restated as follows. Given an underlying/winner/loser form triplet (x, y^*, y) , consider the corresponding vector $\bar{a} = (\bar{a}_1, \dots, \bar{a}_k, \dots, \bar{a}_n)$ of \mathbb{R}^n whose generic k th component \bar{a}_k is defined as the difference between the number of violations $C_k(x, y)$ assigned by constraint C_k to the loser pair (x, y) and the number of violations $C_k(x, y^*)$ assigned by that same constraint to the winner pair (x, y^*) , namely $\bar{a}_k = C_k(x, y) - C_k(x, y^*)$. Then, a triplet (x, y^*, y) and a ranking \gg are OT-compatible iff the vector \bar{a}_{\gg} , obtained by reshuffling the components of \bar{a} according to the linear order \gg , is lexicographically positive. From this perspective, I think that the OT framework was very popular in the operation research literature in the seventies; see for instance Fishburn (1974) for a review.

4.1.2 OT with comparative rows

Given an underlying/winner/loser form triplet (x, y^*, y) , we can classify a given constraint C_k as in (8). In words, WINNER-PREFERRING constraints are those that assign less violations to the winner pair (x, y^*) than to the loser pair (x, y) ; vice versa, LOSER-PREFERRING constraints are those that assign more violations to the winner pair (x, y^*) than to the loser pair (x, y) . Of course, a constraint can be neither winner- nor loser-preferring if it assigns the same number of violations to the two pairs, in which case it is called EVEN. Winner- and loser-preferring constraints together are also called ACTIVE.

- (8) a. Constraint C_k is called WINNER-PREFERRER iff $C_k(x, y^*) < C_k(x, y)$;
 b. constraint C_k is called LOSER-PREFERRER iff $C_k(x, y^*) > C_k(x, y)$;
 c. constraint C_k is called EVEN iff $C_k(x, y^*) = C_k(x, y)$.

Consider again the underlying/winner/loser form triplet (2b). We usually represent the relevant information concerning this triplet in the form of the OT-tableau (9). Crucially, this representation (9) encodes the actual number of constraint violations, as the number of stars in a given cell.

(9)

winner		
		/rad/
		[rat], [rad]
		[rad]
loser		

 \implies

/rad/	F_{pos}	F_{gen}	M
[rat]		*	
[rad]			*

Yet, the definition (4) of OT-compatibility does not really care about the actual number of constraint violations. It only cares about whether a given constraint is winner-preferring or loser-preferring or neither. Thus, let's replace the overabundant representation (9) with the sharper representation (10).

(10)

winner		
		/rad/
		[rat], [rad]
		[rad]
loser		

 \implies

F_{pos}	F_{gen}	M
INACTIVE	LOSER-PREFERRER	WINNER-PREFERRER

Let me abbreviate (10) as in (11) by means of the three symbols W, L and E that stand for winner-preferring, loser-preferring and even or inactive, respectively. Thus, all the information we really need concerning the underlying/winner/loser form triplet (2b) is just this row with entries equal to either L, or E, or W, one for every constraint.

(11)

winner		
		/rad/
		[rat], [rad]
		[rad]
loser		

 \implies

F_{pos}	F_{gen}	M
E	L	W

These considerations lead to the following developments. After Tesar (1995) and Prince (2002), let me say that a COMPARATIVE TABLEAU is a table of the form (12), with n columns (one for every constraint) and an arbitrary number (say m) of rows, whose elements are W's, L's and E's. I will say that the k th column of the tableau CORRESPONDS to the k th constraint C_k . I will denote an arbitrary comparative tableau by \mathbf{A} . I will also write $\mathbf{A} \in \{L, E, W\}^{m \times n}$ to make explicit the number m of rows and the number n of columns. I will call an arbitrary row of \mathbf{A} a COMPARATIVE ROW³; I will denote it by \mathbf{a} ; and I will denote by $(a_1, \dots, a_k, \dots, a_n)$ the n entries $a_k \in \{L, E, W\}$ of the comparative row \mathbf{a} . I will often omit E's for the sake of readability.

³Various alternative names for comparative rows have been used in the literature: Prince (2002) calls them ELEMENTARY RANKING CONDITIONS; Tesar and Smolensky call them MARK DATA PAIRS.

$$(12) \quad \mathbf{A} = \underbrace{\begin{bmatrix} C_1 & \dots & C_k & \dots & C_n \\ W & L & W & L & E \\ L & W & W & E & E \\ E & W & W & L & L \end{bmatrix}}_{n \text{ columns}} \left. \vphantom{\begin{bmatrix} C_1 & \dots & C_k & \dots & C_n \\ W & L & W & L & E \\ L & W & W & E & E \\ E & W & W & L & L \end{bmatrix}} \right\} m \text{ rows}$$

Let me introduce next the new notion of OT-COMPATIBILITY in (13), as a purely combinatoric relation that holds between a ranking and a comparative row or tableau.

- (13) a. A ranking \gg is called OT-COMPATIBLE with a comparative row \mathbf{a} iff the highest \gg -ranked active constraint is a winner-preferring constraint.
- b. A ranking \gg is called OT-COMPATIBLE with a comparative row \mathbf{A} iff it is OT-compatible with each of its rows according to (13a).
- c. A comparative tableau \mathbf{A} is called OT-COMPATIBLE iff \mathbf{A} is OT-compatible with at least one ranking according to (13b).

A ranking \gg is OT-compatible with an underlying/winner/loser form triplet according to the old notion of OT-compatibility (4a) iff \gg is OT-compatible according to the new notion of OT-compatibility (13a) with the comparative row corresponding to that triplet according to (14). Furthermore, a ranking \gg is OT-compatible with a set of underlying/winner/loser form triplets according to the old notion of OT-compatibility (4b) iff \gg is OT-compatible according to the new notion of OT-compatibility (13b) with the comparative tableau obtained by organizing one underneath the other (in any order) the comparative rows corresponding to those triplets according to (14).

$$(14) \quad \begin{array}{c} \text{winner} \\ | \\ (x, y^*, y) \\ | \\ \text{loser} \end{array} \implies \mathbf{a} = [a_1 \quad \dots \quad a_k \quad \dots \quad a_n]$$

$$\text{where } a_k \doteq \begin{cases} W & \text{if } C_k \text{ is winner-preferring} \\ L & \text{if } C_k \text{ is loser-preferring} \\ E & \text{if } C_k \text{ is inactive} \end{cases}$$

Throughout this work, I will assume that the OT on-line algorithm is given as input an OT-compatible comparative tableau. It is therefore useful to have good characterizations of comparative tableaux that are OT-compatible according to (13c). To start, note that OT-compatibility can obviously be made pictorially explicit as follows: a ranking \gg is OT-compatible with a comparative tableau \mathbf{A} iff, once the n columns of \mathbf{A} are reordered from left to right in decreasing order according to \gg , then the leftmost non-E entry of every row of the reordered tableau is a w. To illustrate again with the typology in (1), consider the set of underlying/winner/loser form triplets $\{ (/da/, [da], [ta]), (/rad/, [rat], [rad]) \}$. The corresponding comparative tableau is (15a). There is only one ranking OT-compatible with this set of triplets according to definition (4b), namely the ranking $F_{\text{pos}} \gg M \gg F_{\text{gen}}$ in (3). If the columns of the tableau in (15a) are reordered according to this ranking from left to right in decreasing order, we get the tableau (15b), which is the one in (15a) with the last two columns switched. The tableau (15b) thus obtained has the property that the leftmost non-E symbol of every row is a w. No other ordering of the columns of the tableau in (15a) has this property.

$$(15) \quad \text{a.} \quad \begin{array}{c} \text{winners} \\ | \\ (/da/, [da], [ta]) \\ (/rad/, [rat], [rad]) \\ | \\ \text{losers} \end{array} \implies \begin{bmatrix} F_{\text{pos}} & F_{\text{gen}} & M \\ W & W & L \\ E & L & W \end{bmatrix}$$

$$\text{b.} \quad \begin{array}{c} \text{winners} \\ | \\ (/da/, [da], [ta]) \\ (/rad/, [rat], [rad]) \\ | \\ \text{losers} \end{array} \implies \begin{bmatrix} F_{\text{pos}} & M & F_{\text{gen}} \\ W & L & W \\ E & W & L \end{bmatrix}$$

The following claim 1 slightly sharpens the characterization of OT-compatibility just discussed; a very different characterization will be offered by claim 41 in chapter 8. I adopt here the following piece of notation. Consider a comparative row $\mathbf{a} = (a_1, \dots, a_n)$ and a ranking \gg . Let $dec_{\gg}(\mathbf{a})$ be the \gg -DECISIVE constraint for the comparative row \mathbf{a} , namely the \gg -highest ranked active constraint, as defined in (16a). Definition (13a) thus says that \gg is OT-compatible with the comparative row \mathbf{a} iff the \gg -decisive constraint $dec_{\gg}(\mathbf{a})$ for that row \mathbf{a} is a winner-preferring constraint. Consider a comparative tableau \mathbf{A} and a ranking \gg . For every constraint $k = 1, \dots, n$, let $dec_{\gg}(k)$ be the set of those rows of the tableau \mathbf{A} whose \gg -decisive constraint is C_k , as defined in (16b).

$$(16) \quad \begin{aligned} \text{a. } & dec_{\gg}(\mathbf{a}) \doteq \max_{\gg} \{ C_k \mid a_k \neq E \} \\ \text{b. } & dec_{\gg}(C_k) = \{ \mathbf{a} \mid dec_{\gg}(\mathbf{a}) = C_k \} \end{aligned}$$

If \gg is OT-compatible with \mathbf{A} , then (16b) can be made more explicit as in (17a). As noted above, I can assume without loss of generality that this OT-compatible ranking \gg is $C_1 \gg C_2 \gg \dots \gg C_n$, since the labeling of the constraints is arbitrary to start with. Under this assumption, (17a) becomes (17b). In this case, I will simplify the notation by dropping the subscript \gg , thus writing just $dec(k)$.

$$(17) \quad \begin{aligned} \text{a. } & dec(C_k) = \left\{ \mathbf{a} = (a_1, \dots, a_n) \mid \begin{array}{l} a_k = W \\ a_h = E \text{ for every } C_h \gg C_k \end{array} \right\} \\ \text{b. } & dec(C_k) = \left\{ \mathbf{a} = (a_1, \dots, a_n) \mid \begin{array}{l} a_k = W \\ a_1 = \dots = a_{k-1} = E \end{array} \right\} \end{aligned}$$

Claim 1 A comparative tableau \mathbf{A} is OT-compatible iff there exists an integer $d \leq n$ such that \mathbf{A} can be brought into the form (18) by reordering its rows and its columns and by relabeling the constraints.

$$(18) \quad \begin{array}{c} \begin{array}{cccccc} C_1 & C_2 & \dots & C_d & C_{d+1} & \dots & C_n \end{array} \\ \begin{array}{l} dec(C_1) \\ \\ dec(C_2) \\ \\ \vdots \\ \\ dec(C_d) \end{array} \left[\begin{array}{cccccc} W & & & & & & \\ | & \dots & \dots & \dots & \dots & \dots & \dots \\ W & & & & & & \\ \hline E & W & & & & & \\ | & | & \dots & \dots & \dots & \dots & \dots \\ E & W & & & & & \\ \hline E & E & & & & & \\ \vdots & \vdots & & & & & \\ E & E & & & & & \\ \hline E & E & & W & & & \\ | & E & \dots & | & \dots & \dots & \dots \\ E & E & & W & & & \end{array} \right] \\ \underbrace{\hspace{10em}}_{d \text{ columns}} \quad \underbrace{\hspace{10em}}_{n-d \text{ columns}} \end{array}$$

In words: the rows in the first block start with a w, followed by arbitrary entries; the rows in the second block start with an E and a w, followed by arbitrary entries; and so on, until the rows in the last block that start with $d - 1$ E's and a w, followed by arbitrary entries.

Proof. Let \gg be a ranking OT-compatible with \mathbf{A} . Partition the rows of \mathbf{A} into the sets $dec_{\gg}(C_k)$ for every constraint $k = 1, \dots, n$. Some of these sets $dec_{\gg}(C_k)$ might of course be empty. Let $d \leq n$ be the total number of constraints C_k such that $dec_{\gg}(C_k)$ is non-empty. Relabel the constraints so that these d constraints are C_1, \dots, C_d and that C_1 is the \gg -highest ranked among $\{C_1, \dots, C_d\}$, C_2 is the next \gg -highest ranked among $\{C_1, \dots, C_d\}$, ... C_d is the \gg -bottom ranked among $\{C_1, \dots, C_d\}$. Reorder the columns of \mathbf{A} as follows: let the leftmost column be the one corresponding to C_1 ; let the second left-most column be the one corresponding to C_2 ; ... let the d th left-most column be the one corresponding to C_d ; let the remaining columns be the ones corresponding to the remaining constraints, in any order. Furthermore, reorder the rows of \mathbf{A} as follows: let the top rows be the ones in $dec_{\gg}(C_1)$; let the next top rows be the ones in $dec_{\gg}(C_2)$; and so on. The comparative tableau \mathbf{A} thus reordered has the form (18). ■

To conclude this discussion of OT-compatible comparative tableaux, let me introduce and briefly comment on the equivalence relation over comparative tableaux in (19).

- (19) Two comparative tableaux $\mathbf{A} \in \{\text{L, E, W}\}^{m \times n}$ and $\mathbf{A}' \in \{\text{L, E, W}\}^{m' \times n}$ (with the same number n of columns but possibly different numbers m and m' of rows) are called OT-EQUIVALENT iff any ranking is OT-compatible with \mathbf{A} iff it is OT-compatible with \mathbf{A}' .

There are a number of operations that transform a comparative tableau \mathbf{A} into a new comparative tableau \mathbf{A}' OT-equivalent to it; see Prince (2002) for more discussion. Here, I only discuss two such operations, that will often be useful in what follows. Here is a first operation that obviously preserves OT-compatibility: if a row of a comparative tableau \mathbf{A} contains no entries equal to L, then \mathbf{A} is OT-equivalent to the tableau \mathbf{A}' obtained from \mathbf{A} by suppressing that row. Here is another operation that preserves OT-compatibility: if a row of a comparative tableau \mathbf{A} contains ℓ entries equal to L, then \mathbf{A} is equivalent to the tableau \mathbf{A}' obtained from \mathbf{A} by replacing that row with ℓ rows identical to it but for the fact that each of them retains only one of the original ℓ entries equal to L while the others are replaced by E's. The latter operation is illustrated in (20) w.r.t. the last row of the comparative tableau \mathbf{A} .

$$(20) \quad \mathbf{A} = \begin{bmatrix} \text{W} & \text{L} & & \text{W} \\ & \text{W} & \text{L} & \text{W} \\ \text{W} & \text{W} & \text{L} & \text{L} \end{bmatrix} \implies \mathbf{A}' = \begin{bmatrix} \text{W} & \text{L} & & \text{W} \\ & \text{W} & \text{L} & \text{W} \\ \text{W} & \text{W} & \text{L} & \\ \text{W} & \text{W} & & \text{L} \end{bmatrix}$$

These simple considerations prove claim 2. This claim entails in particular that I can restrict myself without loss of generality to comparative tableaux that have the property that each one of their rows contains one and only one entry equal to L.

Claim 2 For every comparative tableau \mathbf{A} there is another comparative tableau \mathbf{A}' such that \mathbf{A} is OT-equivalent to \mathbf{A}' , has exactly one entry equal to L per row and has a number of rows polynomial in the number of rows and columns of \mathbf{A} .

4.1.3 OT with ranking vectors

So far, I have represented rankings as total orders on the constraint set. As noted in Boersma (1997, 2008), a ranking over n constraints can equivalently be represented as an n -tuple of numbers, exploiting the natural ordering between numbers. To introduce the idea, consider again the three constraints in (1d), repeated in (21).

$$(21) \quad \begin{array}{l} C_1 = F_{\text{pos}} = \text{IDENT}[\text{ONSET}][\text{VOICE}], \\ C_2 = F_{\text{gen}} = \text{IDENT}[\text{VOICE}], \\ C_3 = M = *[\text{+VOICE}, \text{-SONORANT}]. \end{array}$$

Let's pair up each one of these three constraints C_1 , C_2 and C_3 with an arbitrary number, for instance with the three numbers θ_1 , θ_2 and θ_3 in (22).

$$(22) \quad \begin{array}{ccc} C_1 = F_{\text{pos}} & C_2 = F_{\text{gen}} & C_3 = M \\ | & | & | \\ \theta_1 = 100 & \theta_2 = 2 & \theta_3 = 50 \end{array}$$

As noted in (23), this triplet $\theta = (\theta_1, \theta_2, \theta_3)$ intuitively univocally represents the ranking in (3): the positional faithfulness constraint $C_1 = F_{\text{pos}}$ is top ranked, because it corresponds to the largest number $\theta_1 = 100$; the general faithfulness constraint $C_2 = F_{\text{gen}}$ is bottom ranked, because it corresponds to the smallest number $\theta_2 = 2$; the markedness constraint $C_3 = M$ is ranked in between, because its corresponding number $\theta_3 = 50$ lies in between the two numbers corresponding to the two faithfulness constraints.

$$(23) \quad \theta = \begin{pmatrix} F_{\text{pos}} & F_{\text{gen}} & M \\ 100 & 2 & 50 \end{pmatrix} \implies F_{\text{pos}} \gg M \gg F_{\text{gen}}$$

It is crucial for the correspondence in (23) to hold that the three numbers θ_1 , θ_2 and θ_3 are all distinct. What if among these three numbers there are, say, two that are identical? Well, we might think of the corresponding triplet $\theta = (\theta_1, \theta_2, \theta_3)$ as representing two different rankings “at the same time”. An example is provided in (24): the numbers assigned to the markedness constraint M and to the general faithfulness constraint F_{gen} tie; thus, this triplet θ can be interpreted as representing two different rankings at the same time, depending on how the tie between the markedness constraint M and the general faithfulness constraint F_{gen} is resolved. We can think of these two rankings as two different ways of “refining” the tie in the triplet θ .

$$(24) \quad \theta = \begin{pmatrix} F_{\text{pos}} & F_{\text{gen}} & M \\ 100 & 50 & 50 \end{pmatrix} \begin{array}{l} \nearrow F_{\text{pos}} \gg M \gg F_{\text{gen}} \\ \searrow F_{\text{pos}} \gg F_{\text{gen}} \gg M \end{array}$$

These considerations can be generalized as follows. After Boersma (1997), let me say that a RANKING VECTOR is an n -tuple θ of numbers $\theta_1, \dots, \theta_n$ (one for every constraint), as in (25). The k th component θ_k of θ is called the RANKING VALUE of the corresponding constraint C_k .

$$(25) \quad \theta = \begin{pmatrix} C_1 & \dots & C_k & \dots & C_n \\ \theta_1 & \dots & \theta_k & \dots & \theta_n \end{pmatrix}.$$

Given a comparative row $\mathbf{a} = (a_1, \dots, a_n)$, let me denote by $W(\mathbf{a})$ and by $L(\mathbf{a})$ the corresponding sets of winner- and loser-preferring constraints, as in (26). These two sets capture all the relevant information concerning the comparative row \mathbf{a} . The constraints in $W(\mathbf{a}) \cup L(\mathbf{a})$ are the active ones.

$$(26) \quad \begin{aligned} W(\mathbf{a}) &= \left\{ k \in \{1, \dots, n\} \mid a_k = W \right\} = \text{set of winner-preferring constraints} \\ L(\mathbf{a}) &= \left\{ k \in \{1, \dots, n\} \mid a_k = L \right\} = \text{set of loser-preferring constraints} \end{aligned}$$

Let me now introduce the new notion of OT-compatibility in (27), as a combinatoric relation that holds between a comparative row or tableau and a ranking vector.

- (27) a. A ranking vector $\theta = (\theta_1, \dots, \theta_n)$ is called OT-COMPATIBLE with a comparative row \mathbf{a} iff the condition $\boxed{\max_{k \in W(\mathbf{a})} \theta_k > \max_{h \in L(\mathbf{a})} \theta_h}$ holds.
- b. A ranking vector $\theta = (\theta_1, \dots, \theta_n)$ is called OT-COMPATIBLE with a comparative tableau \mathbf{A} iff it is compatible with every row of \mathbf{A} according to (27b).
- c. A comparative tableau \mathbf{A} is called OT-COMPATIBLE iff \mathbf{A} is OT-compatible with at least one ranking vector according to (27b).

Let me illustrate the relationship between the old notion of OT-compatibility in (13) and the new notion in (27). Let me say that a ranking \gg is a REFINEMENT of a ranking vector $\theta = (\theta_1, \dots, \theta_n)$ iff condition (28) holds for every $h, k = 1, \dots, n$. The idea is as follows: if two or more components of θ tie, different refinements of θ can break the tie in different ways; otherwise, any refinement satisfies the ordering implicitly encoded into the natural ordering of the components of θ . The boxed condition in (27a) says that there exists a winner-preferring constraint whose ranking value is larger than the ranking value of any loser-preferring constraint. Thus, a ranking vector θ is OT-compatible with a comparative row \mathbf{a} according to the new definition (27a) iff every refinement of θ is a ranking OT-compatible with \mathbf{a} according to the old definition (13a) of OT-compatibility. Furthermore, a comparative tableau \mathbf{A} is OT-compatible according to the new definition (27c) iff \mathbf{A} is OT-compatible according to the old definition (13c).⁴

⁴ Note that the notion of OT-compatibility (27) for ranking vectors with two or more identical components has nothing to do with the alternative notion of OT-compatibility introduced by Tesar and Smolensky (2000), that allows for multiple constraints to be assigned to the same stratum with the corresponding tie resolved additively. Without getting into the details of this alternative definition of OT-compatibility, let me illustrate the difference with an example. Consider the comparative row in (ia) together with the ranking vector in (ib), with the two identical components $\theta_1 = \theta_2$.

$$(i) \quad \mathbf{a} = \begin{array}{ccc} C_1 & C_2 & C_3 \\ \mathbf{w} & \mathbf{L} & \mathbf{w} \end{array}$$

$$(28) \quad \theta_h > \theta_k \implies C_h \gg C_k.$$

I will denote by $\mathbf{0}$ the null ranking vector, namely the ranking vector whose components are all equal to zero. Given two ranking vectors $\theta = (\theta_1, \dots, \theta_n)$ and $\theta' = (\theta'_1, \dots, \theta'_n)$, I will denote by $\theta \geq \theta'$ the condition that each component of θ is larger than or equal to the corresponding component of θ' , namely $\theta_h \geq \theta'_h$ for every component $h = 1, \dots, n$. I will denote by $\theta = \theta' + \theta''$ the COMPONENT-WISE SUM of the two ranking vectors $\theta' = (\theta'_1, \dots, \theta'_n)$ and $\theta'' = (\theta''_1, \dots, \theta''_n)$, namely the vector $\theta = (\theta_1, \dots, \theta_n)$ defined by $\theta_h = \theta'_h + \theta''_h$ for every component $h = 1, \dots, n$. I will denote by $\theta = \max\{\theta', \theta''\}$ the COMPONENT-WISE MAXIMUM of the two ranking vectors $\theta' = (\theta'_1, \dots, \theta'_n)$ and $\theta'' = (\theta''_1, \dots, \theta''_n)$, namely the vector $\theta = (\theta_1, \dots, \theta_n)$ defined by $\theta_h = \max\{\theta'_h, \theta''_h\}$ for every component $h = 1, \dots, n$. An important issue for what follows is that of operations on ranking vectors that preserve OT-compatibility, namely operations that take two ranking vectors θ' and θ'' both OT-compatible with a given comparative row and yield a new ranking vector θ OT-compatible with that row too. It is easy to check that the operation of point-wise sum between two ranking vectors does not preserve OT-compatibility. Tesar and Smolensky implicitly note that the operation of component-wise maximum does preserve OT-compatibility, as stated in claim 3.

Claim 3 If two ranking vectors θ' and θ'' are OT-compatible with a comparative row \mathbf{a} , then their component-wise maximum $\theta = \max\{\theta', \theta''\}$ is OT-compatible with \mathbf{a} too.

Proof. The proof consists of the chain of inequalities in (29). Here, I have reasoned as follows: in step (a), I have used the definition of the ranking vector θ as the component-wise maximum of the two ranking vectors θ' and θ'' ; in step (b), I have commuted the two maximum operators; in step (c), I have used the hypothesis that both ranking vectors θ' and θ'' are OT-compatible with the comparative row \mathbf{a} , namely satisfy the boxed condition in (27a); in step (d), I have commuted again the two maximum operators; in step (e), I have used again the definition of the ranking vector θ as the component-wise maximum of the two ranking vectors θ' and θ'' .

$$(29) \quad \max_{h \in W(\mathbf{a})} \theta_h \stackrel{(a)}{=} \max_{h \in W(\mathbf{a})} \underbrace{\max\{\theta'_h, \theta''_h\}}_{\theta_h} \\ \stackrel{(b)}{=} \max \left\{ \max_{h \in W(\mathbf{a})} \theta'_h, \max_{h \in W(\mathbf{a})} \theta''_h \right\} \\ \stackrel{(c)}{>} \max \left\{ \max_{k \in L(\mathbf{a})} \theta'_k, \max_{k \in L(\mathbf{a})} \theta''_k \right\} \\ \stackrel{(d)}{=} \max_{k \in L(\mathbf{a})} \underbrace{\max\{\theta'_k, \theta''_k\}}_{\theta_k} \\ \stackrel{(e)}{=} \max_{k \in L(\mathbf{a})} \theta_k$$

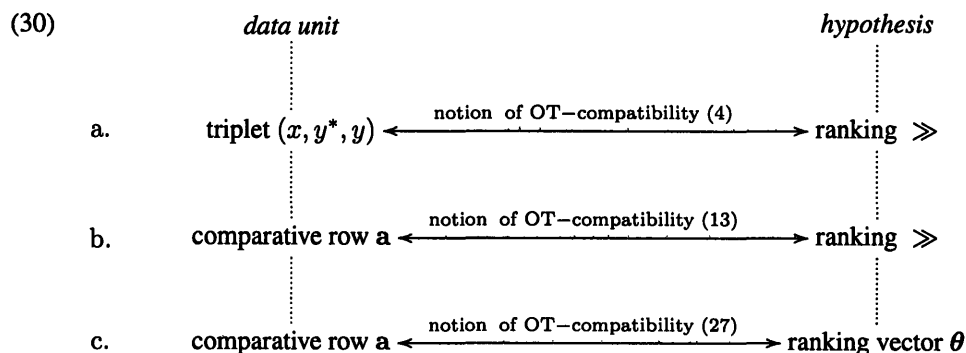
The chain of inequalities (29) shows that the vector θ satisfies the boxed condition in (27a), namely is OT-compatible with the comparative row \mathbf{a} . ■

Let me summarize this section as in (30). Subsection 4.1.1 reviewed the three main ingredients of the OT framework: the basic unit of data, namely underlying/winner/loser form triplets; the basic format of the hypotheses, namely rankings; and the connection between the two, namely the notion

$$\text{b. } \theta = \begin{pmatrix} C_1 & C_2 & C_3 \\ 2 & 2 & 1 \end{pmatrix}$$

According to the alternative definition of OT-compatibility introduced by Tesar and Smolensky (2000), the ranking vector (ib) is indeed OT-compatible with the comparative row (ia), since the two equally highest ranked constraints C_1 and C_2 "cancel out". Instead, the ranking vector (ib) is *not* OT-compatible with the comparative row (ia) according to the definition (27), since the ranking vector (ib) admits the refinement $C_2 \gg C_1 \gg C_3$ that is not OT-compatible with the comparative row (ia) according to the definition (13a). In the rest of the paper, I will stick to the notion of OT-compatibility in (27) and ignore the alternative notion of OT-compatibility introduced by Tesar and Smolensky, that I have just alluded to. The entire discussion is thus framed squarely within standard OT. Contrary to what suggested by Tesar and Smolensky, there is no need to step outside of the standard framework for algorithmic purposes.

of OT-compatibility in (4). Subsection 4.1.2 reviewed an alternative representation of the data in terms of comparative rows and tableaux, and thus restated the notion of OT-compatibility as in (13). Finally, subsection 4.1.3 reviewed an alternative representation of the hypotheses in terms of ranking vectors, and thus restated the notion of OT-compatibility as in (27).



In this subsection, the restatement of OT from rankings to ranking vectors has been presented only as a formal trick, that allows the original roundabout definition of OT-compatibility (13) to be algebraized as the elegant and concise boxed condition in (27). Question (31) now naturally arises.

- (31) Which one of the two parameterizations considered so far, rankings or ranking vectors, is the one actually used indeed by the learner? In other words, does the learner assume that languages are parameterized by *combinatoric* objects such as rankings or by *continuous* objects such as ranking vectors?

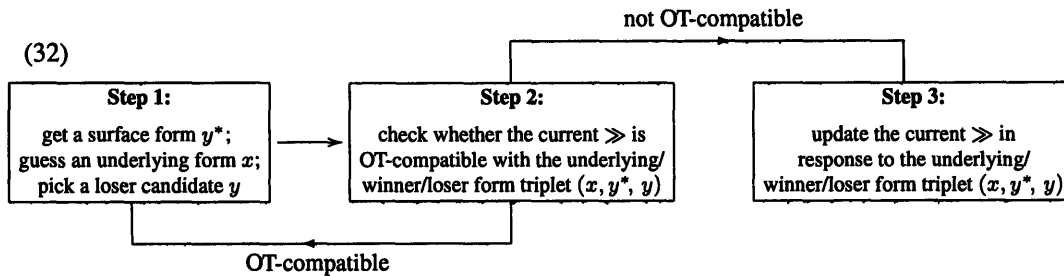
Various cognitive faculties seem to involve continuous parameterizations, such as vision or motor control. If the answer to question (31) turned out to be that natural language learners assume languages to be parameterized by continuous ranking vectors, rather than by combinatoric rankings, then the language faculty would pattern with other cognitive faculties. And the combinatoric nature of linguistics would turn out to be just an illusion. Question (31) is thus an interesting question. Yet, it is a hard question, since the two parameterizations are indistinguishable from many points of view. As discussed below in section 4.3, the technical developments of the OT on-line algorithm seem to point toward the conclusion that OT-grammars should indeed be parameterized by continuous ranking vectors rather than by combinatoric rankings.

4.2 The OT on-line algorithm: a computational perspective

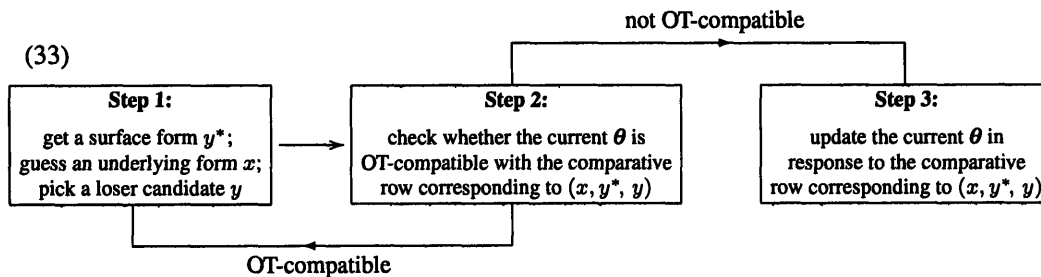
The rest of this chapter describes the simplest possible model for the acquisition of phonology, namely the OT on-line algorithm. It is useful to split up the presentation into its computational and its modeling halves. This section presents the OT on-line algorithm from a purely computational point of view. The next section 4.3 will address various issues that arise in using the algorithm as a model of the acquisition of phonology.

4.2.1 Outline of the algorithm

The general shape of the OT ON-LINE ALGORITHM is described in (32). The algorithm maintains a CURRENT RANKING \gg . This current ranking is initialized to an INITIAL RANKING \gg^{init} . At each time, the algorithm is given a surface form y^* that belongs to the target language. The algorithm figures out a corresponding underlying form x and selects some other candidate surface form y for x different from y^* . If the current ranking vector \gg is OT-compatible with the underlying/winner/loser form triplet (x, y^*, y) according to (4a), then nothing happens and the algorithm loops back. Otherwise, the algorithm takes action by updating its current ranking in response to its failure in accounting for this underlying/winner/loser form triplet (x, y^*, y) .



For future developments, it is useful to restate the model (32) in terms of ranking vectors rather than in terms of rankings, as in (33). The algorithm (33) entertains a CURRENT RANKING VECTOR θ , initialized to a given INITIAL RANKING VECTOR θ^{init} and updated whenever it is not OT-compatible with the comparative row corresponding to the current underlying/winner/loser form triplet.



Given a surface form y^* , step 1 of algorithm (32) asks for a corresponding underlying form x . To determine this underlying form is not at all a trivial task. Here is a possible way to go. Assume that the set of underlying forms \mathcal{X} and the set of surface forms \mathcal{Y} coincide. This assumption makes sense for many realistic cases, although admittedly not for cases such as syllabification, stress assignment, and so on, where the set of output forms is richer than the set of input forms. Under this assumption, a possible solution is (34), namely to assume that the underlying form x corresponding to the given surface form y^* is completely faithful.

(34) In step 1 of algorithm (32), let the underlying form x be defined by $x = y^*$.

Assumption (34) is computationally sound; here is why. Consider a 4-tuple of universal specifications $(\mathcal{X}, \mathcal{Y}, \text{Gen}, \mathcal{C})$ with the crucial property that $\mathcal{X} = \mathcal{Y}$. Following Tesar (2008), let's say that an OT-grammar OT_{\gg} in the corresponding typology corresponding to some ranking \gg is OUTPUT DRIVEN iff it satisfies condition (35) for every candidate surface form $y \in \mathcal{Y}$. This condition says that the surface form y belongs to the language $\mathcal{R}(\text{OT}_{\gg})$ corresponding to the ranking \gg iff the corresponding OT-grammar OT_{\gg} maps that form y (construed as an underlying form) into itself (construed as a surface form). Tesar proves that a generic OT-grammar satisfies condition (35) under mild assumptions on the constraint set. Condition (35) ensures that assumption (34) is sound, in the sense that it cannot possibly lead to a mistake.

$$(35) \quad y \in \mathcal{R}(\text{OT}_{\gg}) \iff \text{OT}_{\gg}(y) = y$$

Having received a surface form y^* and having determined a corresponding underlying form x , say according to (34), step 1 of algorithm (32) asks for a corresponding loser candidate y . A sensible strategy to perform this task is (36). The rationale is that, if the algorithm keeps picking candidates that trigger an update in step 3, then the algorithm will proceed faster toward success.

(36) In step 1 of algorithm (32), pick a loser candidate $y \in \text{Gen}(x)$ such that the current ranking \gg is *not* OT-compatible with the underlying/winner/loser form triplet (x, y^*, y) , if any such candidate y exists.

Of course, there exists a candidate $y \in \text{Gen}(x)$ such that the current ranking \gg is not OT-compatible with the underlying/winner/loser form triplet (x, y^*, y) iff the corresponding grammar OT_{\gg} maps the underlying form x to a surface form $y = \text{OT}_{\gg}(x)$ different from y^* . This suggests that a possible implementation of (36) is (37). Whether this is an efficient implementation or not, it depends on a variety of factors. One of these factors is how long it takes to actually compute $y = \text{OT}_{\gg}(x)$.

- (37) In step 1 of algorithm (32), let the candidate y be defined by $y = \text{OT}_{\gg}(x)$, where \gg is the current ranking.

The restatement (33) of the OT on-line model in terms of ranking vectors raises a small problem for assumption (37); let me illustrate. Assume that the current ranking vector θ at a given iteration of algorithm (33) happens to have pairwise distinct components. We can then consider the OT-grammar OT_{\gg} corresponding to the unique ranking \gg represented by θ . And we can compute the corresponding predicted winner $y \doteq \text{OT}_{\gg}(x)$. Assumption (37) can thus be restated as in (38).

- (38) In step 1 of algorithm (33), let the candidate y be defined by $y \doteq \text{OT}_{\gg}(x)$, where \gg is the ranking univocally represented by the current ranking vector θ .

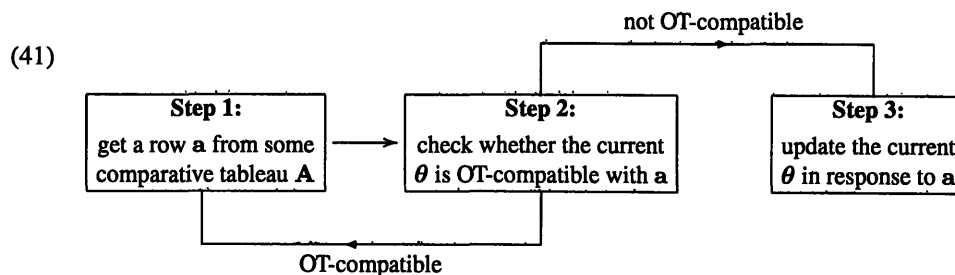
Of course, in the general case the current ranking vector θ might happen to have some identical components and thus does not identify a unique OT grammar. We can of course still run the strategy in (38), by considering the OT-grammar corresponding to an arbitrary refinement of θ , as stated in the variant (39).

- (39) In step 1 of algorithm (33), let the candidate y be defined by $y \doteq \text{OT}_{\gg}(x)$, where \gg is an arbitrary refinement of the current ranking vector θ .

Boersma (1997) suggests a drastic way to implement (39), namely (40). The idea is to add to the current ranking vector $\theta = (\theta_1, \dots, \theta_n)$ a noise vector $\epsilon = (\epsilon_1, \dots, \epsilon_n)$ whose components are sampled independently according to some continuous distribution (say, the uniform distribution $\mathcal{U}([-\sigma, \sigma])$ on some small interval $[-\sigma, \sigma]$ or the normal distribution $\mathcal{N}(0, \sigma)$ with zero mean and variance σ). Indeed, if the components $\epsilon_1, \dots, \epsilon_n$ are sampled *independently* accordingly to a *continuous* distribution, then the probability that $\theta + \epsilon$ has two identical components is null.

- (40) In step 1 of algorithm (33), let the candidate y be defined by $y \doteq \text{OT}_{\gg}(x)$, where \gg is the unique refinement of the vector $\theta + \epsilon$, where $\epsilon = (\epsilon_1, \dots, \epsilon_n)$ i.i.d. according to a proper continuous distribution.

The choice between these various alternatives might have modeling consequences. But it does not have theoretical computational consequences. In fact, all these choices do not affect the worst-case number of updates in step 3, which is the measure of theoretical significance.⁵ Thus, I can abstract away from the issue of the proper definition of step 1, namely from the issue of the proper choice of the underlying form x and of the loser form y . Throughout this section, I thus assume that they are given too. In other words, I simplify the algorithm (33) as in (41). According to (41), the algorithm is given a comparative row a at each iteration; it checks whether the current ranking vector θ is OT-compatible with that comparative row a ; if it is, nothing happens and the algorithm gets another comparative row; if it is not, then the algorithm takes action by updating its current ranking vector in response to its failure in accounting for the current comparative row a . I assume that the rows fed to the algorithm in step 1 are sampled from a fixed, given OT-compatible comparative tableau A , called the INPUT to the algorithm.



To complete the description of the OT on-line model, we need to discuss update rules to be used in step 3. An UPDATE RULE for step 3 of algorithm (41) is a function of the form (42), that takes a comparative row a together with the current ranking vector θ^{old} and returns a new updated ranking vector θ^{new} .

⁵Boersma (1997) and Boersma and Hayes (2001) actually suggest that this variant (40) does turn out to be useful to handle certain cases of variation.

(42) update rule : $(\theta^{\text{old}}, \mathbf{a}) \mapsto \theta^{\text{new}}$

Most of the update rules considered in the literature as well as in this work satisfy the two conditions in (43). Assumption (43a) says that there is no vacuous update: when update is performed, something has got to change in the current ranking vector. Assumption (43b) entails in particular that update rules are not sensitive to the actual size of the components of the current ranking vector θ^{old} . This last assumption will be violated only in subsections 4.2.4, 5.5.2 and 5.5.3.

- (43) a. The updated ranking vector θ^{new} is different from the current ranking vector θ^{old} whenever the current vector θ^{old} is not OT-compatible with the current comparative row \mathbf{a} .
- b. The dependence of the update rule on the current ranking vector θ^{old} is rather limited, in the sense that the update rule does not see the difference between two current ranking vectors that admit the same refinements, namely that represent the same rankings.

Let me say that the on-line algorithm (41), with a specific update rule (42) used in step 3, CONVERGES iff condition (44) holds. This condition ensures that, if every row of the input comparative tableau is supplied to the algorithm a sufficient number of times, then the algorithm will eventually entertain a current ranking vector OT-compatible with that comparative tableau. No further update will therefore happen. Let me call such vector the FINAL ranking vector in the run considered, and denote it by θ^{fin} .

- (44) If the comparative rows in step 1 are sampled from an OT-compatible input tableau, then the algorithm can perform only a finite number of updates in step 3 (no matter the choice of the initial ranking vector and the order that the comparative rows are fed with).

To develop a theory of an update rule (42) for the OT on-line algorithm (41) means at least to answer the three questions listed in (45).⁶ This section systematizes within the unifying framework of ranking vectors various results of the form (45) for a number of update rules that have been developed in the literature within slightly different frameworks (standard OT with total hierarchies, OT with non-total hierarchies, stochastic OT, etcetera).

- (45) a. Does the on-line algorithm (41) corresponding to a given update rule (42) converge in finite time, namely does it satisfy condition (44)?
- b. If it does, what is the worst case number of updates in step 3 for an arbitrary input comparative tableau \mathbf{A} with n columns and an arbitrary initial vector θ^{init} ?
- c. And how can the final vector θ^{fin} be characterized in terms of the input comparative tableau \mathbf{A} and the initial ranking vector θ^{init} ?

Let me close this introduction with a synopsis of the main results reviewed from the literature in the rest of this section. Given a comparative row \mathbf{a} , recall the definition (26) of the sets $W(\mathbf{a})$ and $L(\mathbf{a})$ of winner- and loser-preferring constraints. The various update rules considered in the literature fit into the general scheme (46a), described in words in (46b). Let me unpack the notation. The ranking vector $\theta^{\text{new}} = (\theta_1^{\text{new}}, \dots, \theta_n^{\text{new}})$, obtained by updating the current ranking vector $\theta^{\text{old}} = (\theta_1^{\text{old}}, \dots, \theta_n^{\text{old}})$ according to (46) in response to a comparative row \mathbf{a} , is the component-wise sum of the current ranking vector θ^{old} and a numerical vector $\bar{\mathbf{a}} = (\bar{a}_1, \dots, \bar{a}_k) \in \mathbb{R}^n$. This vector $\bar{\mathbf{a}}$ has the following shape: if C_k is a winner-preferring constraint, then \bar{a}_k is null or positive, so that update increases the ranking value of C_k (or leaves it unchanged); if C_k is a loser-preferring constraint, then \bar{a}_k is null or negative, so that update decreases the ranking value of C_k (or leaves it unchanged); if C_k is not active, then \bar{a}_k is null, so that update leaves the ranking value of C_k unchanged. This vector $\bar{\mathbf{a}}$ is called the UPDATE VECTOR corresponding to θ^{old} and \mathbf{a} . The positive components of the update vector are called PROMOTION AMOUNTS; the (absolute value of the) negative components of the update vector are called DEMOTION AMOUNTS. To define a specific update rule of the form (46) means to define the update vector $\bar{\mathbf{a}}$, as a function of the current ranking vector θ^{old} and the current comparative row \mathbf{a} . The notation $\bar{\mathbf{a}}$ is not optimal, because it hides the dependence of the update

⁶A fourth interesting question concerns robustness to noise, namely the performance of the OT on-line algorithm (41) when the input comparative tableau is *not* OT-compatible. But I ignore this issue here, and always assume that the input tableau is indeed OT-compatible.

vector on the current ranking vector θ^{old} . Yet, assumption (46b), that an update rule is only sensitive to the rankings represented by the current ranking vector θ^{old} and not to the current ranking vector itself, entails in particular that the dependence of the update vector on the current ranking vector is limited, while the crucial dependence is the one on the current comparative row. In particular, it means that fixed the comparative row \mathbf{a} , there is only a finite number of corresponding update vectors. All update rules considered in this work will have the additive form in (46), but for the one considered in subsection 5.5.3.

$$(46) \quad \text{a.} \quad \begin{bmatrix} \theta_1^{\text{new}} \\ \vdots \\ \theta_k^{\text{new}} \\ \vdots \\ \theta_n^{\text{new}} \end{bmatrix} = \begin{bmatrix} \theta_1^{\text{old}} \\ \vdots \\ \theta_k^{\text{old}} \\ \vdots \\ \theta_n^{\text{old}} \end{bmatrix} + \begin{bmatrix} \bar{a}_1 \\ \vdots \\ \bar{a}_k \\ \vdots \\ \bar{a}_n \end{bmatrix} \quad \text{where } \bar{a}_k = \begin{cases} \geq 0 & \text{if } k \in W(\mathbf{a}) \\ \leq 0 & \text{if } k \in L(\mathbf{a}) \\ 0 & \text{otherwise} \end{cases}$$

b. If the constraint C_k is winner-preferring (loser-preferring), add to its current ranking value a corresponding null or positive (null or negative) constant \bar{a}_k .

The various update rules considered in the literature differ along three dimensions, that can be preliminarily sketched as follows. The *first dimension* is whether the rule is demotion-only or promotion-demotion, depending on whether the promotion amounts are always null or not. The *second dimension* is whether the rule is minimal or non-minimal, depending on whether it only demotes the loser-preferring constraints that ‘need’ to be demoted or all of them. The *third dimension* is whether the rule is gradual or non-gradual, depending on whether a given comparative row can trigger two consecutive updates or not. I classify in (47) the main update rules considered so far in the literature according to these three dimensions. I also provide the current name of the corresponding OT on-line algorithm (41) together with known results for the two properties (45a) and (45b) concerning convergence and speed.

$$(47) \quad \text{update rules} \left\{ \begin{array}{l} \text{demotion-only} \\ \text{demotion-promotion} \end{array} \right. \left\{ \begin{array}{l} \text{gradual} \\ \text{non-gradual} \end{array} \right. \left\{ \begin{array}{l} \text{minimal} \\ \text{non-minimal} \end{array} \right. \begin{array}{l} \Rightarrow \text{GLA}_{\text{min}}^{\text{dem.}} \\ \text{converges, and is fast} \\ \Rightarrow \text{GLA}^{\text{dem.}} \\ \text{converges, but is slow} \\ \Rightarrow \text{CD:} \\ \text{converges, and is super-fast} \\ \Rightarrow \text{GLA}_{\text{min}} \\ \text{does not converge} \\ \Rightarrow \text{GLA:} \\ \text{does not converge} \\ \Rightarrow ?? \end{array}$$

Tesar (1995), Tesar (1998), Tesar and Smolensky (1998) and Tesar and Smolensky (2000, Ch. 7) (henceforth: Tesar and Smolensky) offer a complete theory (45) for minimal demotion-only update rules. Subsections 4.2.2-4.2.4 illustrate how the most natural framework to set this theory is that of ranking vectors.⁷ Subsection 4.2.5 shows that non-minimal update rules are inferior to minimal update rules, and should thus be avoided. Some promotion-demotion update rules were tried out by Boersma (1997). Yet, they were shown not to converge by a recent counterexample due to Pater (2008). Thus, the problem of devising convergent promotion-demotion update rules is currently open in the literature. Section 4.2.6 provides a detailed explanation of Pater’s counterexample. This explanation previews the use of dual variables in the analysis of the on-line algorithm (41). Dual variables will become crucial in chapter 5, where I will take on the task of devising provably convergent promotion-demotion update rules.

⁷The idea of representing rankings in terms of numerical ranking vectors is actually implicitly already present in Tesar and Smolensky’s notion of the OFFSET of a constraint w.r.t. a ranking, defined as the number of strata above that constraint in that ranking. I think indeed that the most natural framework to develop their theory is the one adopted here, in terms of ranking vectors, whereby most of their reasoning can be straightforwardly algebraized.

4.2.2 Generalities on demotion-only update rules

The best studied class of update rules for step 3 of the on-line algorithm (41) is that of DEMOTION-ONLY update rules. These are update rules of the form (46) whose promotion-amounts are always null. More explicitly, these are rules that update the current ranking vector $\theta^{\text{old}} = (\theta_1^{\text{old}}, \dots, \theta_n^{\text{old}})$ to the new ranking vector $\theta^{\text{new}} = (\theta_1^{\text{new}}, \dots, \theta_n^{\text{new}})$ in response to a comparative row \mathbf{a} according to the scheme in (48a), described in words in (48b).

(48) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} :

$$\text{a. } \begin{bmatrix} \theta_1^{\text{new}} \\ \vdots \\ \theta_k^{\text{new}} \\ \vdots \\ \theta_n^{\text{new}} \end{bmatrix} = \begin{bmatrix} \theta_1^{\text{old}} \\ \vdots \\ \theta_k^{\text{old}} \\ \vdots \\ \theta_n^{\text{new}} \end{bmatrix} + \begin{bmatrix} \bar{a}_1 \\ \vdots \\ \bar{a}_k \\ \vdots \\ \bar{a}_n \end{bmatrix} \quad \text{where } \bar{a}_k = \begin{cases} \leq 0 & \text{if } k \in L(\mathbf{a}) \\ 0 & \text{otherwise} \end{cases}$$

b. Only the ranking values of loser-preferring constraints are modified; whenever a ranking value is modified, it is decreased.

Tesar and Smolensky’s claim 4 ensures that the convergence condition (44) holds for any demotion-only update rule. The core idea of the proof is that demotion-only update rules yield a very simple dynamics over time of the components of the current ranking vector. Simple in the sense that a single component can be straightforwardly analyzed independently of the other components, as long as the components are analyzed “in the right order”, namely in the order provided by a ranking OT-compatible with the input comparative tableau.

Claim 4 The on-line algorithm (41) run with any demotion-only update rule (48), starting from any initial vector, and with comparative rows sampled from an an OT-compatible input comparative tableau can perform only a finite number of updates in step 3.

Proof. Let me prove the claim by induction on the number n of columns of the input comparative tableau \mathbf{A} . The claim trivially holds for $n = 1$. Assume that the claim holds for $n - 1$. Let me prove that the claim holds for n . Consider an OT-compatible input comparative tableau \mathbf{A} with n columns. By claim 1, I can assume without loss of generality that this tableau \mathbf{A} has the shape in (49), for some comparative sub-tableau \mathbf{A}' with $n - 1$ columns.

$$(49) \quad \text{dec}(C_1) \quad \left[\begin{array}{c|ccc} C_1 & C_2 & \dots & C_n \\ \hline W & & \dots & \\ \vdots & & & \\ W & & \dots & \\ \hline E & & & \\ \vdots & & & \mathbf{A}' \\ E & & & \end{array} \right]$$

The ranking value θ_1 of constraint C_1 will stay put at its initial value, because its column does not contain a single L and thus its ranking value never gets modified by a demotion-only update rule. Each row in the upper block can only trigger a finite number of updates, i.e. at most as many updates as needed to bring underneath C_1 every constraint that is loser-preferring for some row in the top block.⁸ Thus, there is a finite time after which the algorithm will only respond to rows in the bottom block of the tableau. Thus, from that time on, the algorithm behaves as if it was given as input the comparative tableau \mathbf{A}' . Since the tableau \mathbf{A}' has only $n - 1$ columns, the claim follows by the inductive hypothesis. ■

⁸This claim is not true in the general case. For instance, it would not be true if the demotion amounts were allowed to become arbitrarily small over time. But this possibility is ruled out by the assumption (43b) that the update vector only depends on the current comparative row and on the rankings represented by the current ranking vector. Thus in particular, each comparative row is paired up only with a finite number of update vectors, so that the demotion amounts cannot become arbitrarily small.

Claim 4 guarantees a positive answer to question (45a) for any demotion-only update rule. By (48), different demotion-only update rules differ along two dimensions: by which loser-preferring constraints get demoted and by how much they get demoted. In the next three subsections, I consider various specific instances of demotion-only update rules that differ along these two dimensions. For some of these specific demotion-only update rules, we will be able to tackle questions (45b) and (45c) as well.

4.2.3 First example of demotion-only update rule: minimal and gradual

Given a ranking vector $\theta = (\theta_1, \dots, \theta_n)$, let $L(\mathbf{a}, \theta)$ be the subset (50) of those loser-preferring constraints $C_k \in L(\mathbf{a})$ that are ranked above the top ranked winner-preferring constraint, namely the set of those loser-preferring constraints C_k whose ranking value θ_k is not smaller than the largest ranking value $\max_{h \in W(\mathbf{a})} \theta_h$ of winner-preferring constraints. Tesar and Smolensky call any such loser-preferring constraint UNDOMINATED w.r.t. θ . An update rule is called MINIMAL if it only demotes currently undominated loser-preferring constraints, namely $\theta_k^{\text{old}} > \theta_k^{\text{new}}$ entails that $k \in L(\mathbf{a}, \theta^{\text{old}})$.

$$(50) \quad L(\mathbf{a}, \theta) \doteq \left\{ k \in \{1, \dots, n\} \mid a_k = L, \theta_k \geq \max_{h \in W(\mathbf{a})} \theta_h \right\}$$

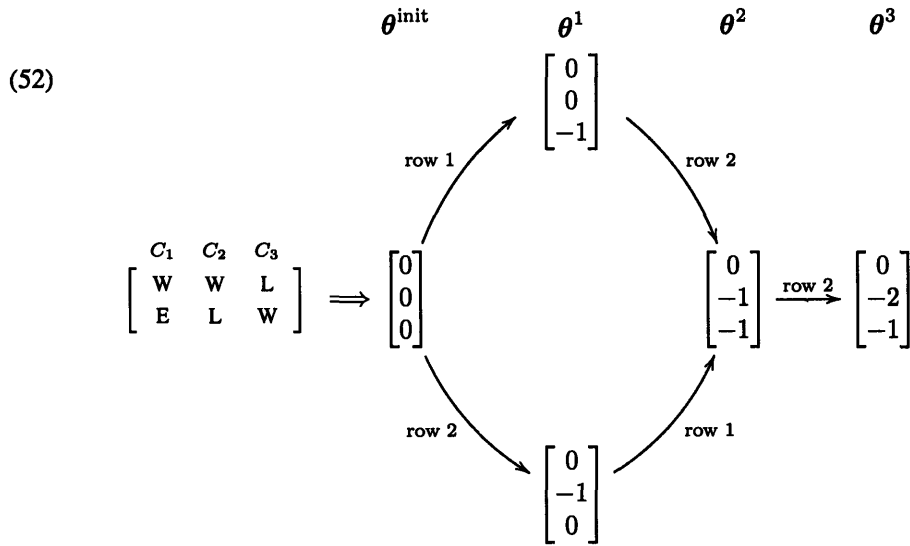
Consider the rule that updates the current ranking vector $\theta^{\text{old}} = (\theta_1^{\text{old}}, \dots, \theta_n^{\text{old}})$ to the new ranking vector $\theta^{\text{new}} = (\theta_1^{\text{new}}, \dots, \theta_n^{\text{new}})$ in response to a comparative row \mathbf{a} as in (51a), described in words in (51b). This update rule is demotion-only and minimal: it demotes currently undominated loser-preferring constraints by a small fixed amount. According to (51a), the fixed small amount by which the currently undominated loser-preferring constraints get demoted is 1. This amount 1 can of course be replaced by any positive constant $\eta > 0$. This constant is called either the STEP-SIZE or the PLASTICITY of the update rule. The on-line algorithm (41) with this minimal, gradual, demotion-only update rule is Boersma's (1997) MINIMAL DEMOTION-ONLY GRADUAL LEARNING ALGORITHM (henceforth: $\text{GLA}_{\text{min}}^{\text{dem}}$).

(51) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} :

$$\text{a. } \theta_k^{\text{new}} = \begin{cases} \theta_k^{\text{old}} - 1 & \text{if } k \in L(\mathbf{a}, \theta^{\text{old}}) \\ \theta_k^{\text{old}} & \text{otherwise} \end{cases}$$

b. Demote by 1 the currently undominated loser-preferring constraints.

The behavior of the on-line algorithm (41) with the update rule (51) run on the input comparative tableau (15) starting from the null initial vector is described by the diagram (52). At the first iteration, the algorithm can receive either the first or the second row of the comparative tableau. Suppose that it receives the first row. Since the current ranking vector θ^{init} is not OT-compatible with that row, then the algorithm updates θ^{init} to θ^1 by demoting by 1 the ranking value of constraint C_3 , which is the only constraint that has an L in the first comparative row. At the next iteration, the algorithm can again receive either the first or the second row of the comparative tableau. Since the current ranking vector θ^1 is OT-compatible with the first row, nothing happens if the algorithm receives the first row. Since the current ranking vector θ^1 is not OT-compatible with the second row, once the algorithm receives the second row, it updates its current ranking vector θ^1 to θ^2 by demoting by 1 the ranking value of C_2 , which is the only constraint that has an L in the second comparative row. And so on. No matter the order that the comparative rows are fed to the algorithm, after three updates the algorithm entertains the ranking vector $\theta^3 = (0, -2, -1)$, that represents the ranking $C_1 \gg C_3 \gg C_2$. Since this ranking vector is OT-compatible with the input tableau, no further update will be triggered. This ranking vector is thus the final vector entertained by the algorithm.



Boersma (1998, p. 323-327) notes that the analysis of Tesar and Smolensky (originally developed for the update rule presented in the next subsection) trivially extends to the case of the update rule (51). In the rest of this subsection, I present Tesar and Smolensky’s analysis in full detail.

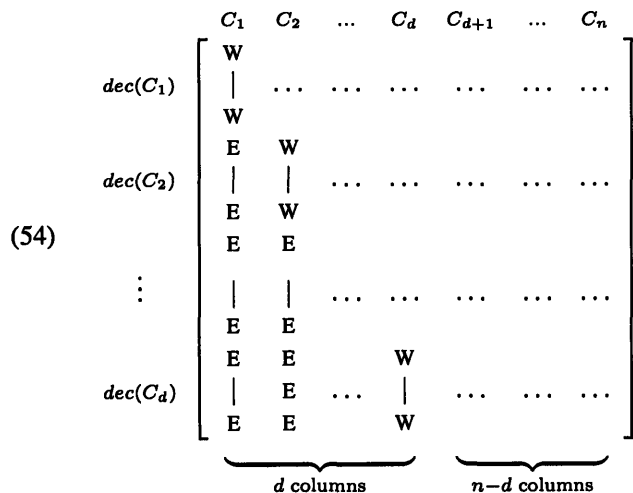
4.2.3.1 Bound on the worst-case number of updates

The input comparative tableau in (52) is only OT-compatible with the ranking that assigns C_1 to the first stratum, C_3 to the second stratum and C_2 to the third stratum. The diagram in (52) shows that the ranking value of C_1 never goes below 0 throughout the entire learning path, the ranking value of C_3 never goes below 1 and the ranking value of C_2 never goes below 2. In other words, the ranking value of a constraint assigned to the k th stratum (with the 1st stratum being the top stratum) never goes below $k - 1$. Tesar and Smolensky note that this property holds in general, as stated in claim 5. In words, this claim says that the components of the current ranking vector cannot become too small. The proof is basically identical to the proof of the preceding general claim 4.

Claim 5 Without loss of generality, assume that the ranking OT-compatible with the input comparative tableau A is $C_1 \gg C_2 \gg \dots \gg C_n$. Then, the ranking vector $\theta = (\theta_1, \dots, \theta_n)$ entertained at a generic time by the on-line algorithm (41) run on the input tableau A with the update rule (51) starting from the null initial vector satisfies condition (53) for every $k = 1, \dots, n$.

(53) $\theta_k \geq -(k - 1)$

Proof. By claim 1, I can assume that the input tableau has the shape in (18), repeated in (54).



Here is why (53) holds for the case $k = 1$. The column of \mathbf{A} corresponding to C_1 does not contain a single L. Thus, the demotion-only update rule (51) never modifies the initial ranking $\theta_1^{\text{init}} = 0$ of constraint C_1 . Here is why (53) holds for the case $k = 2$. The only rows that can trigger demotion of C_2 are the rows in $\text{dec}(C_1)$ that have an L corresponding to C_2 that is currently ranked higher than the W corresponding to C_1 . But once θ_2 reaches -1 , these rows cannot trigger any further update of the ranking value of C_2 . In fact, $\theta_1 = 0$ and thus C_2 is no longer undominated, since the winner-preferring constraint C_1 is ranked above it. In conclusion, the ranking value of C_2 can never get below $\theta_2 = -1$. The cases $k > 2$ are dealt with analogously. ■

I will note in subsection 4.2.5 that claim 5 does not hold for non-minimal update rules. I will note in chapter 5 that a slight generalization of claim 5 actually holds for *any* minimal update rule of the form (46), no matter whether it is demotion-only or not. As shown by the proof of the following claim 6, the invariant (53) immediately yields a bound on the worst case number of updates, thus answering question (45b) for the case of the minimal demotion-only update rule (51).

Claim 6 The on-line algorithm (41) with the update rule (51) run on an arbitrary comparative tableau \mathbf{A} with n columns starting from the null initial vector can perform at most $\frac{1}{2}n(n-1)$ updates in step 3.

Proof. By claim 4, we know that the algorithm does indeed converge to a ranking vector $\theta^{\text{fin}} = (\theta_1^{\text{fin}}, \dots, \theta_n^{\text{fin}})$ OT-compatible with the input tableau \mathbf{A} after a finite number T of updates. Without loss of generality, assume that the input tableau \mathbf{A} is OT-compatible with the ranking $C_1 \gg C_2 \gg \dots \gg C_n$. The proof of the claim consists of the chain of inequalities in (55).

$$(55) \quad T \stackrel{(a)}{\leq} \sum_{k=1}^n -\theta_k^{\text{fin}} \\ \stackrel{(b)}{\leq} \sum_{k=1}^n (k-1) \\ = \frac{1}{2}n(n-1)$$

Here, I have reasoned as follows: in step (a), I have noted that each time update is performed, one or more components of the current ranking vector get decreased by one; since all components start out null, then the total number of updates T must be smaller than or equal to the sum of the components of the final ranking vector θ^{fin} , multiplied by -1 ; in step (b), I have used the invariant (53), that holds for every ranking vector entertained by the algorithm and thus in particular also for the final one. ■

The bound $\frac{1}{2}n(n-1)$ on the worst-case number of updates provided by claim 6 is tight. In fact, consider comparative tableaux of the form (56), that Riggle (2007) calls DIAGONAL. If the rows are fed in the fixed order $\mathbf{a}_{n-1} \rightarrow \mathbf{a}_{n-2} \rightarrow \dots \rightarrow \mathbf{a}_1$, then exactly $\frac{1}{2}n(n-1)$ updates are required.

$$(56) \quad \begin{array}{c} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_{n-2} \\ \mathbf{a}_{n-1} \end{array} \left[\begin{array}{cccccc} C_1 & C_2 & C_3 & C_4 & \dots & C_{n-1} & C_n \\ \text{W} & \text{L} & & & & & \\ & \text{W} & \text{L} & & & & \\ & & \text{W} & \text{L} & & & \\ & & & \ddots & & & \\ & & & & \ddots & & \\ & & & & & \text{L} & \\ & & & & & \text{W} & \text{L} \end{array} \right]$$

To see this, consider for instance the case $n = 4$. As shown by the learning path (57b) described by the algorithm when the rows of the diagonal tableau (57a) are fed in the fixed order $\mathbf{a}_1 \rightarrow \mathbf{a}_2 \rightarrow \mathbf{a}_3$, it takes $6 = \frac{1}{2}4(4-1)$ updates to reach the unique ranking vector OT-compatible with the input comparative tableau.

$$(57) \quad \text{a. } \begin{matrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{matrix} \begin{bmatrix} \text{W} & \text{L} & & \\ & \text{W} & \text{L} & \\ & & \text{W} & \text{L} \end{bmatrix}$$

$$\text{b. } \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \xrightarrow{\mathbf{a}_1} \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} \xrightarrow{\mathbf{a}_2} \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \end{bmatrix} \xrightarrow{\mathbf{a}_3} \begin{bmatrix} 0 \\ -1 \\ -1 \\ -1 \end{bmatrix} \xrightarrow{\mathbf{a}_2} \begin{bmatrix} 0 \\ -1 \\ -2 \\ -1 \end{bmatrix} \xrightarrow{\mathbf{a}_3} \begin{bmatrix} 0 \\ -1 \\ -2 \\ -2 \end{bmatrix} \xrightarrow{\mathbf{a}_3} \begin{bmatrix} 0 \\ -1 \\ -2 \\ -3 \end{bmatrix}$$

Claim 6 provides the worst case number of updates over *all* possible input comparative tableaux with n columns and *all* possible ways of feeding the rows of that tableau to the algorithm. We might also be interested in the worst case number of updates for a *fixed* input comparative tableaux over *all* possible ways of feeding the rows of that tableau to the algorithm. This worst case number of updates measures the complexity of a given comparative tableau \mathbf{A} . I will come back to this issue at the end of the next paragraph 4.2.3.2.

4.2.3.2 Characterization of the final vector

The invariant stated in claim 5 can be slightly strengthened as in claim 7, by considering not just the set of all possible rankings OT-compatible with the input tableau but rather the set of all possible (integer and non-positive) ranking vectors compatible with it.⁹ The proof of this claim 7 is conceptually identical to the proof of the preceding claim 5.

Claim 7 Let θ be the ranking vector entertained at a generic time by the on-line algorithm (41) run on an input OT-compatible comparative tableau \mathbf{A} with the update rule (51) starting from the null initial vector. The inequality (58) holds for every ranking vector $\mathbf{v} \in \{0, -1, -2, \dots\}^n$ with nonpositive integer components OT-compatible with the input tableau.

$$(58) \quad \theta \geq \mathbf{v}$$

Proof. Consider an arbitrary ranking vector $\mathbf{v} = (v_1, \dots, v_n) \in \{0, -1, -2, \dots\}^n$ OT-compatible with the input tableau \mathbf{A} and let me prove that (59) holds for every $k = 1, \dots, n$. Let \gg be a refinement of \mathbf{v} . Without loss of generality, assume that \gg is $C_1 \gg C_2 \gg \dots \gg C_n$. Let me prove (59) by induction on k .

$$(59) \quad \theta_k \geq v_k$$

The inequality (59) holds for $k = 1$. In fact, since the input comparative tableau \mathbf{A} is OT-compatible with a ranking that assigns C_1 to the top stratum, then the column of \mathbf{A} corresponding to C_1 cannot contain a single L and thus the demotion-only update rule (51) will never modify the initial ranking $\theta_1 = 0$ of constraint C_1 . Let me now prove that (59) holds for a generic k , using the inductive hypothesis that it holds for each h such that $h < k$. In other words, let me prove that, if the k th component of the current ranking vector $\theta^{\text{old}} = (\theta_1^{\text{old}}, \dots, \theta_k^{\text{old}}, \dots, \theta_n^{\text{old}})$ is $\theta_k^{\text{old}} = v_k$, then that component will never be further updated. In other words, let me prove that $k \notin L(\mathbf{a}, \theta^{\text{old}})$ for every row $\mathbf{a} = (a_1, \dots, a_h, \dots, a_k, \dots, a_n)$ of the input comparative tableau. If $a_k \neq \text{L}$, then the constraint C_k is not even a loser-preferred constraint and thus a fortiori $k \notin L(\mathbf{a}, \theta^{\text{old}})$. If $a_k = \text{L}$, then the chain of implications in (60) shows that the loser-preferred constraint C_k is not currently undominated and thus $k \notin L(\mathbf{a}, \theta^{\text{old}})$.

⁹Consider for instance the comparative tableau \mathbf{A} in (ia) and the ranking vector θ in (ib).

$$(i) \quad \text{a. } \mathbf{A} = \begin{bmatrix} C_1 & C_2 & C_3 \\ \text{W} & & \text{L} \\ & \text{w} & \text{L} \end{bmatrix}$$

$$\text{b. } \theta = (1, 1, 3)$$

Since θ satisfies the invariant (53) but does not satisfy the invariant (58), then claim 7 does indeed strengthen claim 5.

- (60) \mathbf{A} is OT compatible with $\mathbf{v} \implies$
- $\xRightarrow{(a)}$ there exists $C_h \in W(\mathbf{a})$ such that $v_h > v_k$
 - $\xRightarrow{(b)}$ there exists $C_h \in W(\mathbf{a})$ such that $v_h > v_k$ and $h \in \{1, \dots, k-1\}$
 - $\xRightarrow{(c)}$ there exists $C_h \in W(\mathbf{a})$ such that $v_h > \theta_k^{\text{old}}$ and $h \in \{1, \dots, k-1\}$
 - $\xRightarrow{(d)}$ there exists $C_h \in W(\mathbf{a})$ such that $\theta_h^{\text{old}} > \theta_k^{\text{old}}$
 - $\xRightarrow{(e)}$ $k \notin L(\mathbf{a}, \theta^{\text{old}})$

In (60), I have reasoned as follows: in step (a), I have noted that, since \mathbf{a} is OT-compatible with the ranking vector \mathbf{v} and since $a_k = L$, then there has got to exist $C_h \in W(\mathbf{a})$ with $v_h > v_k$; in step (b), I have noted that, since $C_1 \gg C_2 \gg \dots \gg C_n$ is a refinement of \mathbf{v} and since $v_h > v_k$, then it must be $h < k$; in step (c), I have used the hypothesis that $\theta_k^{\text{old}} = v_k$; in step (d), I have used the inductive hypothesis that $\theta_h^{\text{old}} \geq v_h$ for every $h = 1, \dots, k-1$; in step (e), I have used definition (50) of $L(\mathbf{a}, \theta^{\text{old}})$. ■

As shown by the proof of the following claim 8, the strengthened invariant (58) immediately yields a complete characterization of the final vector entertained by the algorithm in terms of the input comparative tableau, thus answering question (45c) for the case of the minimal demotion-only update rule (51). Note that the identity (61) makes sense by virtue of claim 3, that ensures that the component-wise maximum of OT-compatible vectors is OT-compatible too.¹⁰

Claim 8 The final ranking vector θ^{fin} returned by the on-line algorithm (41) run with the update rule (51) on an input OT-compatible comparative tableau \mathbf{A} starting from the null initial vector is uniquely characterized by (61), namely as the component-wise maximum over ranking vectors OT-compatible with \mathbf{A} with non-positive integer components.

$$(61) \quad \theta^{\text{fin}} = \max \underbrace{\left\{ \mathbf{v} \in \{0, -1, -2, \dots\}^n \mid \mathbf{v} \text{ is OT-compatible with } \mathbf{A} \right\}}_{(*)}$$

Proof. The final vector θ^{fin} returned by the algorithm is OT-compatible with the input comparative tableau and has integer non-positive components. Thus, θ^{fin} belongs to the set (*) in (61). Thus, to prove (61) it is sufficient to prove that $\theta^{\text{fin}} \geq \mathbf{v}$ holds for every ranking vector \mathbf{v} in the set (*) in (61). This fact is guaranteed by the invariant (58). ■

The ranking vector $\theta^{\text{fin}} = (\theta_1^{\text{fin}}, \dots, \theta_n^{\text{fin}})$ defined in (61) only depends on the input comparative tableau \mathbf{A} . The worst case number of updates over all possible ways of feeding the rows of that input tableau \mathbf{A} is thus bounded by $-\sum_{h=1}^n \theta_h^{\text{fin}}$. The bound is tight in the case of comparative tableaux that contain a unique entry equal to L per row; the bound is not tight for comparative rows that contain multiple entries equal to L in some rows.

4.2.3.3 Extension to an arbitrary initial vector

So far, I have only considered the case where the initial ranking vector θ^{init} is the null vector. The analysis trivially extends to the case of any CONSTANT initial vector, namely the case of an initial

¹⁰Strictly speaking, claim 3 does not apply to the case in (61), since the set (*) in (61) is not finite. But this is only a small technicality, that can be immediately overcome as follows. Consider an arbitrary ranking vector $\bar{\mathbf{v}}$ in the set (*) in (61). Thus, the identity (i) holds: the component-wise maximum over the set (*) is identical to the component-wise maximum over the set (**) in (i).

$$(i) \quad \max \underbrace{\left\{ \mathbf{v} \in \{0, -1, \dots\}^n \mid \mathbf{v} \text{ OT-comp. with } \mathbf{A} \right\}}_{(*)} = \max \underbrace{\left\{ \mathbf{v} \in \{0, -1, \dots\}^n \mid \mathbf{v} \text{ OT-comp. with } \mathbf{A}, \mathbf{v} \geq \bar{\mathbf{v}} \right\}}_{(**)}$$

Since the set (**) is finite, then claim 3 does apply, and thus ensures that its component-wise maximum is OT-compatible with the tableau \mathbf{A} .

vector that has identical (although possibly nonnull) components. Furthermore, the analysis also easily extends to the case of a general initial vector θ^{init} , yielding claim 9. The proof of this claim is a straightforward variant of the preceding analyses; I provide it here explicitly because the claim has not so far appeared in the literature.

Claim 9 The final ranking vector θ^{fin} returned by the on-line algorithm (41) run with the update rule (51) on an input OT-compatible comparative tableau \mathbf{A} starting from an arbitrary integral initial vector $\theta^{\text{init}} = (\theta_1^{\text{init}}, \dots, \theta_n^{\text{init}})$ is uniquely characterized by (62), namely as the component-wise maximum over ranking vectors OT-compatible with \mathbf{A} obtained from the initial vector θ^{init} by component-wise sum with a non-positive integral vector \mathbf{v} .

$$(62) \quad \theta^{\text{fin}} = \max \underbrace{\left\{ \theta = \theta^{\text{init}} + \mathbf{v} \mid \mathbf{v} \in \{0, -1, -2, \dots\}^n, \theta \text{ OT-compatible with } \mathbf{A} \right\}}_{(*)}$$

Furthermore, the algorithm can perform at most $\Delta(\theta^{\text{init}}) + \frac{1}{2}n(n-1)$ updates in step 3, where $\Delta(\theta^{\text{init}})$ is the quantity defined as follows:

$$(63) \quad \Delta(\theta^{\text{init}}) = \sum_{h=1}^n \left(\theta_h^{\text{init}} - \min_{h=1} \theta_h^{\text{init}} \right)$$

and thus measures the “scatteredness” of the initial vector θ^{init} .

Proof. The proof of the identity (62) is identical to the proof of the identity (61). Let me prove that the number of updates is upper bounded by $\Delta(\theta^{\text{init}}) + \frac{1}{2}n(n-1)$. Without loss of generality, assume that the input tableau \mathbf{A} is OT-compatible with the ranking $C_1 \gg C_2 \gg \dots \gg C_n$. Consider the ranking vector $\theta = (\theta_1, \dots, \theta_n)$ defined as in (64) for every $k = 1, \dots, n$. Note that $\theta_k > \theta_{k+1}$ for every $k = 1, \dots, n-1$, and thus θ is OT-compatible with the tableau \mathbf{A} . Furthermore, the number v_k defined in (64) is integral and nonpositive. Thus, θ belongs to the set (*) in (62).

$$(64) \quad \theta_k = \theta_k^{\text{init}} + \underbrace{\left(\min_{h=1, \dots, n} \theta_h^{\text{init}} - \theta_k^{\text{init}} \right)}_{v_k} - (k-1)$$

The number T of updates performed by the algorithm in a given run can then be upper bounded by means of the chain of inequalities in (65).

$$(65) \quad \begin{aligned} T &\stackrel{(a)}{\leq} \sum_{k=1}^n \left(\theta_k^{\text{init}} - \theta_k^{\text{fin}} \right) \\ &\stackrel{(b)}{\leq} \sum_{k=1}^n \theta_k^{\text{init}} - \sum_{k=1}^n \theta_k \\ &\stackrel{(c)}{=} \sum_{k=1}^n \theta_k^{\text{init}} - \sum_{k=1}^n \left[\theta_k^{\text{init}} + \left(\min_{h=1, \dots, n} \theta_h^{\text{init}} - \theta_k^{\text{init}} \right) - (k-1) \right] \\ &= \sum_{k=1}^n \left(\theta_k^{\text{init}} - \min_{h=1, \dots, n} \theta_h^{\text{init}} \right) + \sum_{k=1}^n (k-1) \\ &= \Delta(\theta^{\text{init}}) + \frac{1}{2}n(n-1) \end{aligned}$$

In step (a), I have reasoned as follows: the k th component of the current ranking vector starts out at θ_k^{init} , gets demoted by 1 each time it is updated and ends up at θ_k^{fin} ; thus, the k th component is updated a number of times that is equal to $\theta_k^{\text{init}} - \theta_k^{\text{fin}}$; the total number of times T that update is performed cannot be larger than the total number of times that the first component of the current ranking vector is updated plus the total number of times that the second component is updated and so on (it can be smaller, because more than one component can be updated in a single update). In step (b), I have used the identity (62) together with the fact that the vector θ defined in (64) belongs to the set (*) in (62). In step (c), I have used the definition (64) of the vector θ . ■

The bound on the worst-case number of updates provided by claim 9 is tight, as shown by the same example in (57a) with the rows fed in the same fix order $a_1 \rightarrow a_2 \rightarrow a_3$ and with an initial ranking vector such as $\theta^{\text{init}} = (4, 3, 2, 1)$.

4.2.4 Second example of demotion-only update rule: minimal and non-gradual

An update rule is called NON-GRADUAL iff the following condition holds: if θ^{old} is updated to θ^{new} in response to the comparative row \mathbf{a} , then θ^{new} is immediately OT-compatible with that row \mathbf{a} . Equivalently, iff no comparative row can trigger two consecutive updates. Otherwise, an update rule is called GRADUAL. The update rule (51) just considered is gradual: in the top path of diagram (52), the second row of the input comparative tableau triggers two consecutive updates. As an example of non-gradual update rules, consider the rule that updates the current ranking vector $\theta^{\text{old}} = (\theta_1^{\text{old}}, \dots, \theta_n^{\text{old}})$ to the new ranking vector $\theta^{\text{new}} = (\theta_1^{\text{new}}, \dots, \theta_n^{\text{new}})$ in response to a comparative row \mathbf{a} as in (66a), described in words in (66b). The on-line algorithm (41) with this update rule is Tesar and Smolensky's CONSTRAINT DEMOTION (henceforth: CD).

(66) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} :

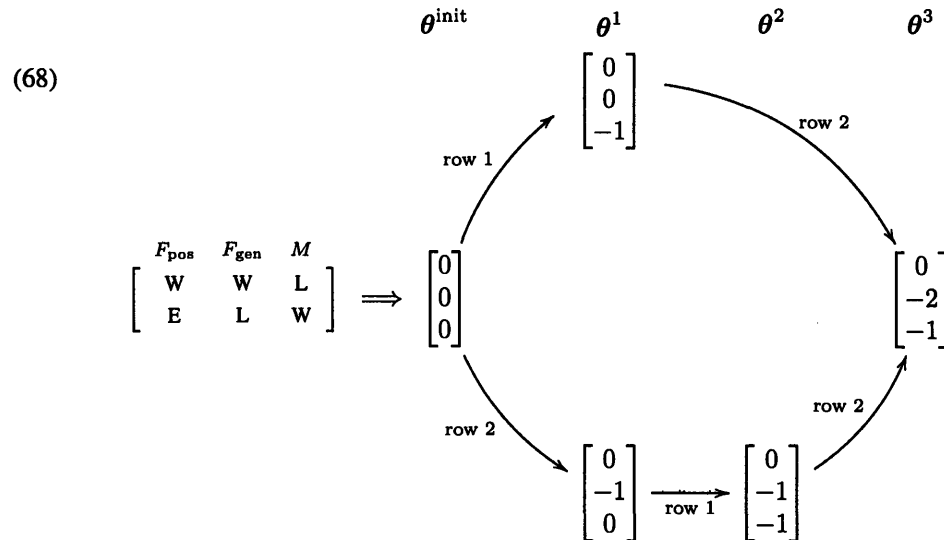
$$\text{a. } \theta_k^{\text{new}} = \begin{cases} \max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} - 1 & \text{if } k \in L(\mathbf{a}, \theta^{\text{old}}) \\ \theta_k^{\text{old}} & \text{otherwise} \end{cases}$$

b. Demote the currently undominated loser-prefering constraints all the way immediately below the currently highest ranked winner-prefering constraint.

This update rule is again *demotion-only*, namely it fits into the scheme (48): in fact, it can be rewritten as in (67), where the demotion amount λ_k is nonnegative because $k \in L(\mathbf{a}, \theta^{\text{old}})$ and thus θ_k^{old} is larger than the ranking value $\max_{h \in W(\mathbf{a})} \theta_h^{\text{old}}$ of the top ranked winner-prefering constraint. Furthermore, this update rule is trivially non-gradual, since all currently undominated loser-prefering constraints are immediately brought to a safe position below the highest ranked winner-prefering constraint.

$$(67) \quad \theta_k^{\text{new}} = \theta_k^{\text{old}} - \underbrace{\left(\theta_k^{\text{old}} - \max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} + 1 \right)}_{\lambda_k}$$

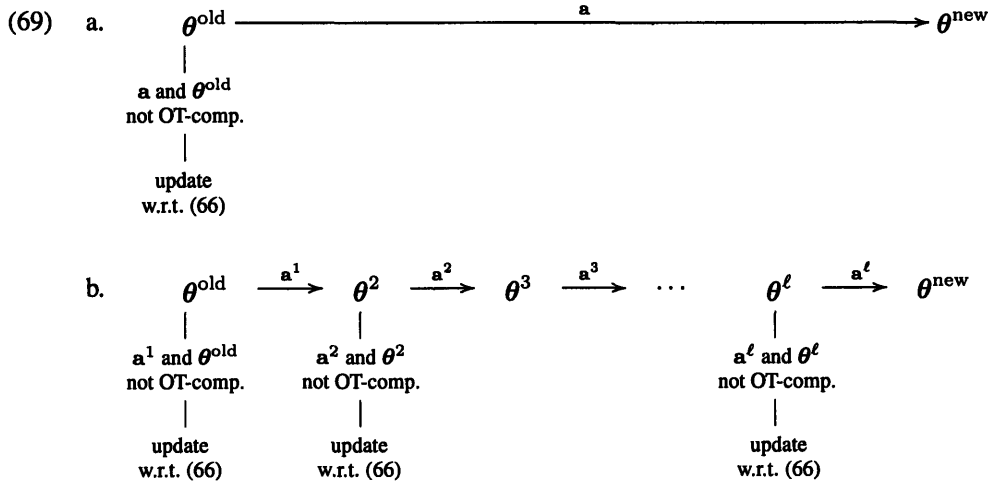
The behavior of the on-line algorithm (41) with the update rule (66) run on the input comparative tableau (15) starting from the null initial vector is described in (68). Note that no row triggers two *consecutive* updates (even though the same row can trigger multiple updates in the same path, as is the case of the second row in the bottom path).



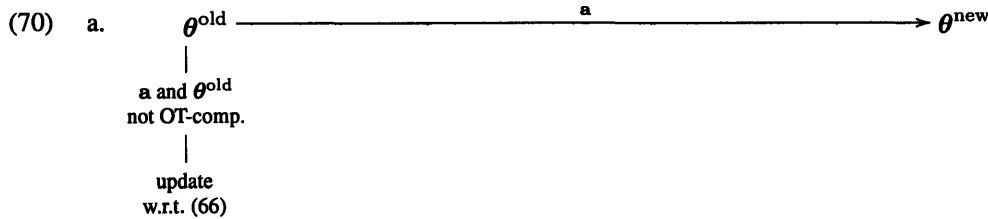
The two diagrams in (52) and (68) only differ because of the fact that the two consecutive updates by the second row in the top path in (52) are replaced by a single “jump” in the case of (68). In other words, the update rule (51) is just a slightly slowed down version of the update rule (66): a single update by the latter corresponds to a series of updates by the former. This trivial idea is used in the following proof of the claim 10 that the preceding analysis of the gradual update rule (51) carries over to the non-gradual update rule (66).¹¹

Claim 10 Claims 6 and 8 also hold for the on-line algorithm (41) with the update rule (66).

Proof. The proof rests on two very simple ideas. Here is the *first* idea. Consider a comparative row a not OT-compatible with the current ranking vector θ^{old} . Let ℓ be the number of currently undominated loser-preferring constraints in a , namely the cardinality of the set $L(a, \theta^{\text{old}})$. Consider the ℓ comparative rows a^1, \dots, a^ℓ identical to a but for the fact that each of them retains only one of these undominated ℓ entries equal to L, while the others get replaced by an E. Then, the update (69a) of θ^{old} triggered by a according to the update rule (66) is equivalent to the sequence of ℓ updates in (69b) triggered by the ℓ rows a^1, \dots, a^ℓ according to that same update rule (66).

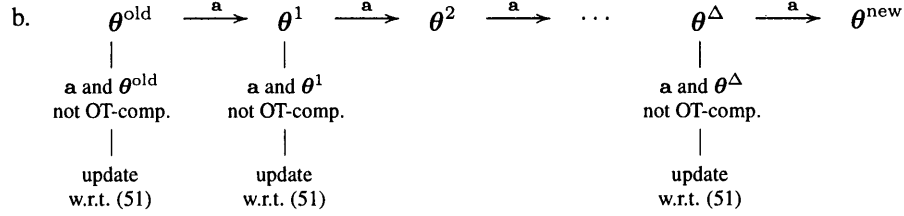


Here is the *second* idea. Consider a comparative row a not OT-compatible with the current ranking vector θ^{old} and assume that a has a unique undominated entry equal to L, corresponding to some constraint C_k . Let Δ be the difference between the current ranking value of C_k and the ranking value of the currently highest ranked winner-preferring constraint, namely $\Delta = \theta_k - \max_{h \in W(a)} \theta_h^{\text{old}}$.¹² Then, the update (70a) of θ^{old} triggered by a according to the non-gradual update rule (66) is equivalent to the sequence of $\Delta + 1$ updates in (70b) triggered by that same row according to the gradual update rule (51).



¹¹Alternatively, I could have repeated *verbatim* for the case of the non-gradual update rule (66) the proofs presented in subsection 4.2.3 for the case of the gradual update rule (51).

¹²Note that Δ is positive, because θ^{old} is not OT-compatible with a , namely $\theta_k^{\text{old}} > \max_{h \in W(a)} \theta_h$.



The claim thus follows from the fact that any update w.r.t. the non-gradual update rule (66) can be mimicked by a series of updates w.r.t. the gradual update rule (51), by combining (69) and (70). ■

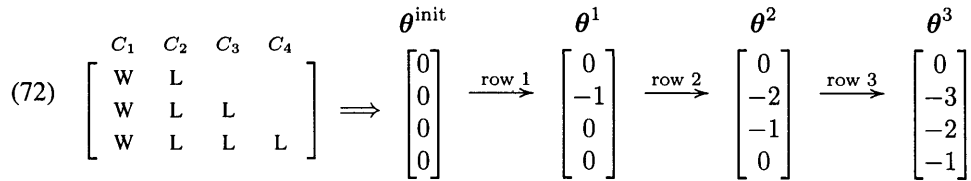
The example in (57) shows that the bound $\frac{1}{2}n(n - 1)$ on the worst case number of updates is tight also for the case of the non-gradual update rule (66). The only difference between the gradual update rule (51) and the non-gradual version (66) shows up in the case of an arbitrary initial vector. As seen in claim 9, the worst case number of updates required by the gradual update rule (51) starting from the initial vector θ^{init} depends on the “scatteredness” of θ^{init} through the quantity $\Delta(\theta^{\text{init}})$ defined in (63). As noted by Tesar and Smolensky, the worst case number of updates required by the non-gradual update rule (66) is instead $n(n - 1)$, and thus does not depend on the initial vector. The latter bound is tight, as shown again by the same example in (57), with an initial ranking vector such as $\theta^{\text{init}} = (4, 3, 2, 1)$.

4.2.5 Third example of demotion-only update rule: non-minimal and gradual

Consider the rule that updates the current ranking vector $\theta^{\text{old}} = (\theta_1^{\text{old}}, \dots, \theta_n^{\text{old}})$ to the new ranking vector $\theta^{\text{new}} = (\theta_1^{\text{new}}, \dots, \theta_n^{\text{new}})$ in response to a comparative row \mathbf{a} as in (71a), described in words in (71b). This update rule is of course again demotion-only, as the two preceding rules (51) and (66); and it is of course again gradual, as the preceding update rule (51). Contrary to the two preceding rules (51) and (66), this new update rule (71) is *non-minimal*, because it demotes all loser-preferring constraints rather than just the currently undominated ones, namely because the set $L(\mathbf{a}, \theta^{\text{old}})$ has been replaced by $L(\mathbf{a})$. The on-line algorithm (41) with this maximal, gradual, demotion-only update rule is Boersma’s (1997) DEMOTION-ONLY GRADUAL LEARNING ALGORITHM. (henceforth: GLA^{dem}).

- (71) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} :
 - a. $\theta_k^{\text{new}} \leftarrow \begin{cases} \theta_k^{\text{old}} - 1 & \text{if } k \in L(\mathbf{a}) \\ \theta_k^{\text{old}} & \text{otherwise} \end{cases}$
 - b. Demote by 1 all loser-preferring constraints.

The invariant (53), that holds for the two minimal update rules (51) and (66), does not hold for the non-minimal update rule (71). The intuitive reason is that a given L can be demoted for “unsubstantial” reasons. This is illustrated by the run in (72): the ranking value of the constraint C_2 drops down to -3 , even though the input comparative tableau is OT-compatible with the ranking $C_1 \gg C_2 \gg C_3 \gg C_4$, where C_2 is assigned to the second stratum.



I want to suggest that non-minimal update rules are “wrong”, in the sense that they ignore the intrinsic logic of OT. Here is a way to make this point. As noted in claim 2, a given comparative tableau \mathbf{A} is OT-equivalent to the comparative tableau \mathbf{A}' obtained from \mathbf{A} as follows: a row of \mathbf{A} that contains ℓ entries equal to L is replaced in \mathbf{A}' with ℓ rows identical to it but for the fact that each of them retains only one of the original ℓ entries equal to L while the others are replaced by E’s. The latter operation is illustrated in (73) w.r.t. the comparative tableau in (72).

$$(73) \quad \mathbf{A} = \begin{bmatrix} \text{W} & \text{L} & & & \\ \text{W} & \text{L} & \text{L} & & \\ \text{W} & \text{L} & \text{L} & \text{L} & \end{bmatrix} \implies \mathbf{A}' = \begin{bmatrix} \text{W} & \text{L} & & & \\ \text{W} & \text{L} & & & \\ \text{W} & & \text{L} & & \\ \text{W} & \text{L} & & & \\ \text{W} & & \text{L} & & \\ \text{W} & & & & \text{L} \end{bmatrix}$$

The crucial difference between minimal and non-minimal update rules can now be spelled out as follows. Minimal update rules match the intrinsic logic of OT, by ensuring that the behavior of the corresponding on-line algorithm (41) does not “see” the difference between a given input comparative tableau \mathbf{A} and the corresponding OT-equivalent tableau \mathbf{A}' obtained by splitting rows with multiple entries equal to L. This point was made explicit above in the proof of claim 10. Non-minimal update rules instead do not match the intrinsic logic of OT, since the corresponding on-line algorithm (41) might behave differently, say, in the case of the two comparative tableaux \mathbf{A} and \mathbf{A}' in (73), despite the fact that they are OT-equivalent. One side effect of the fact that non-minimal update rules are “wrong” is that they might require a number of updates much larger than minimal updates rule. Claim 11 makes this point dramatically explicit. The proof consists of a simple generalization of the case in (72).

Claim 11 The on-line algorithm (41) with the non-minimal update rule (71) has worst case running time super-polynomial in the number n of columns of the input comparative tableau.

Proof. For every $h = 1, \dots, n - 1$ and for every $k = 1, \dots, n - h$, consider the comparative row \mathbf{a}_k^h defined as in (74): it has a W as its h th entry, followed by k L’s, while all other entries are E’s.

$$(74) \quad \mathbf{a}_k^h = \left[\dots \quad C_h \quad C_{h+1} \quad \dots \quad C_{h+k} \quad \dots \right]$$

To illustrate, I give in (75) the comparative tableau obtained by assembling all such rows in the case $n = 6$.

$$(75) \quad \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 \\ \mathbf{a}_1^1 & \text{W} & \text{L} & & & & \\ \mathbf{a}_2^1 & \text{W} & \text{L} & \text{L} & & & \\ \mathbf{a}_3^1 & \text{W} & \text{L} & \text{L} & \text{L} & & \\ \mathbf{a}_4^1 & \text{W} & \text{L} & \text{L} & \text{L} & \text{L} & \\ \mathbf{a}_5^1 & \text{W} & \text{L} & \text{L} & \text{L} & \text{L} & \text{L} \\ \mathbf{a}_1^2 & & \text{W} & \text{L} & & & \\ \mathbf{a}_2^2 & & \text{W} & \text{L} & \text{L} & & \\ \mathbf{a}_3^2 & & \text{W} & \text{L} & \text{L} & \text{L} & \\ \mathbf{a}_4^2 & & \text{W} & \text{L} & \text{L} & \text{L} & \text{L} \\ \mathbf{a}_1^3 & & & \text{W} & \text{L} & & \\ \mathbf{a}_2^3 & & & \text{W} & \text{L} & \text{L} & \\ \mathbf{a}_3^3 & & & \text{W} & \text{L} & \text{L} & \text{L} \\ \mathbf{a}_1^4 & & & & \text{W} & \text{L} & \\ \mathbf{a}_2^4 & & & & \text{W} & \text{L} & \text{L} \\ \mathbf{a}_1^5 & & & & & \text{W} & \text{L} \end{matrix}$$

Suppose that we feed these rows to the algorithm as follows. We consider these rows one at the time, in the order described in (76), namely first all the rows \mathbf{a}_k^h with $h = 1$ ordered from $k = 1$ to $k = n - 1$; then all the rows \mathbf{a}_k^h with $h = 2$ ordered from $k = 1$ to $k = n - 2$; and so on. We keep feeding each row until the current ranking vector is OT-compatible with it, and then we move to the next row. Note that, once the current ranking vector is OT-compatible with a given row \mathbf{a}_k^h , any subsequent ranking vector entertained by the algorithm will remain OT-compatible with that row.

$$\begin{aligned}
(76) \quad & \mathbf{a}_1^1 \Rightarrow \mathbf{a}_2^1 \Rightarrow \dots \Rightarrow \mathbf{a}_{n-2}^1 \Rightarrow \mathbf{a}_{n-1}^1 \Rightarrow \\
& \Rightarrow \mathbf{a}_1^2 \Rightarrow \mathbf{a}_2^2 \Rightarrow \dots \Rightarrow \mathbf{a}_{n-2}^2 \Rightarrow \\
& \vdots \\
& \Rightarrow \mathbf{a}_1^{n-2} \Rightarrow \mathbf{a}_2^{n-2} \\
& \Rightarrow \mathbf{a}_1^{n-1}
\end{aligned}$$

It can easily be seen that the number of updates required by the algorithm to converge to the unique ranking vector OT-compatible with the input comparative tableau is superpolynomial in n when the rows are fed in the order described. ■

4.2.6 Promotion-demotion update rules

Another class of update rules to be used in step 3 of the on-line algorithm (41) that have been studied in the literature is that of PROMOTION-DEMOTION update rules. These are update rules of the form (46) whose promotion amounts are nonnull. More explicitly, these are rules that update the current ranking vector $\theta^{\text{old}} = (\theta_1^{\text{old}}, \dots, \theta_n^{\text{old}})$ to the new ranking vector $\theta^{\text{new}} = (\theta_1^{\text{new}}, \dots, \theta_n^{\text{new}})$ in response to a comparative row \mathbf{a} according to the scheme (77a), described in words in (77b).

(77) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} :

$$\text{a. } \begin{bmatrix} \theta_1^{\text{new}} \\ \vdots \\ \theta_k^{\text{new}} \\ \vdots \\ \theta_n^{\text{new}} \end{bmatrix} = \begin{bmatrix} \theta_1^{\text{old}} \\ \vdots \\ \theta_k^{\text{old}} \\ \vdots \\ \theta_n^{\text{old}} \end{bmatrix} + \begin{bmatrix} \bar{a}_1 \\ \vdots \\ \bar{a}_k \\ \vdots \\ \bar{a}_n \end{bmatrix} \quad \text{where } \bar{a}_k \text{ is } \begin{cases} \leq 0 & \text{if } k \in L(\mathbf{a}) \\ > 0 & \text{if } k \in W(\mathbf{a}) \\ = 0 & \text{otherwise} \end{cases}$$

b. The ranking value of loser-preferring constraints is decreased or left unchanged; the ranking value of winner-preferring constraints is increased.

As an example, consider the rule that updates the current ranking vector $\theta^{\text{old}} = (\theta_1^{\text{old}}, \dots, \theta_n^{\text{old}})$ to the new ranking vector $\theta^{\text{new}} = (\theta_1^{\text{new}}, \dots, \theta_n^{\text{new}})$ in response to a comparative row \mathbf{a} as in (78a), described in words in (78b). This promotion-demotion update rule is *non-minimal*, because it targets all loser-preferring constraints. And it is *gradual*, because the current ranking values are modified only by a small amount, namely 1. The on-line algorithm (41) with this update rule is Boersma's (1997) (deterministic) GRADUAL LEARNING ALGORITHM (henceforth: GLA).

(78) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} :

$$\text{a. } \theta_k^{\text{new}} = \begin{cases} \theta_k^{\text{old}} - 1 & \text{if } k \in L(\mathbf{a}) \\ \theta_k^{\text{old}} + 1 & \text{if } k \in W(\mathbf{a}) \\ \theta_k^{\text{old}} & \text{otherwise} \end{cases}$$

b. Decrease by 1 the ranking values of all loser-preferring constraints; increase by 1 the ranking values of all winner-preferring constraints.

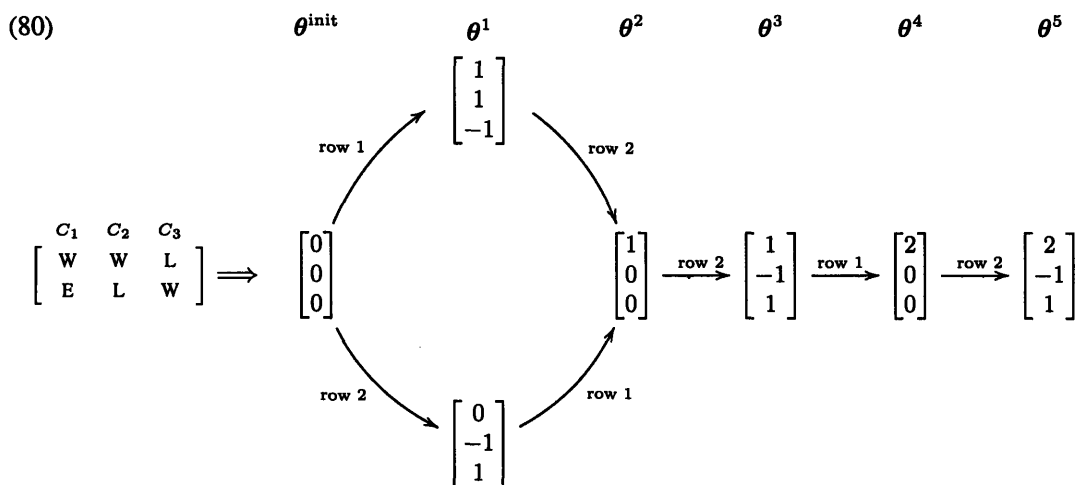
As discussed in subsection 4.2.5, non-minimal update rules do not look like a good option, since a given constraint might get dragged down for irrelevant reasons and thus the resulting algorithm might need many more updates before it converges to a ranking vector OT-compatible with the input tableau. From this perspective, it is thus natural to consider the variant of (78) stated in (79). The promotion-demotion update rule (79) is minimal, since it only demotes the currently undominated loser-preferring constraints. The on-line algorithm (41) with this update rule is Boersma's (1997) MINIMAL GRADUAL LEARNING ALGORITHM (henceforth: GLA_{min}).

(79) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} :

$$\text{a. } \theta_k^{\text{new}} = \begin{cases} \theta_k^{\text{old}} - 1 & \text{if } k \in L(\mathbf{a}, \theta^{\text{old}}) \\ \theta_k^{\text{old}} + 1 & \text{if } k \in W(\mathbf{a}) \\ \theta_k^{\text{old}} & \text{otherwise} \end{cases}$$

- b. Decrease by 1 the ranking value of *undominated* loser-preferring constraints; increase by 1 the ranking values of all winner-preferring constraints.

The behavior of the on-line algorithm (41) with either of these two update rules (78) or (79) on the input tableau (15) starting from the null vector is described in the diagram (80).



Boersma's promotion-demotion update rules (78) and (79) are *symmetric*, in the sense that the demotion amount is equal to the promotion amount (in this case they are both equal to 1, although the actual amount of course does not matter). In the rest of this section, we will see that symmetric promotion-demotion update rules do not work in the general case. In chapter 5, I will thus turn to the investigation of *asymmetric* promotion-demotion update rules, namely update rules where the promotion amount and the demotion amount are different.

4.2.6.1 A detailed explanation of Pater's counterexample

As noted by Tesar and Smolensky (1998, pp. 244-245) in the celebrated passage quoted in (81), promotion-demotion update rules are much harder to analyze than demotion-only update rules.

- (81) “[The update rules considered in the preceding sections 4.2.2-4.2.5 are] defined entirely in terms of demotion; all movement of constraints is downward in the hierarchy. One could reasonably ask if this is an arbitrary choice; couldn't the learner just as easily promote constraints toward the correct hierarchy? The answer is no, and understanding why reveals the logic behind [demotion-only update rules]. [In order for OT-compatibility to hold], at least one [winner-preferring constraint] must dominate all [loser-preferring constraints]. Demotion moves the [loser-preferring constraints]. [...] Once the highest-ranked [winner-preferring constraint] is identified, all of the [loser-preferring constraints] need to be dominated by it, so all [loser-preferring constraints] are demoted if not already so dominated. A hypothetical promotion operation would move the constraints corresponding to the [winner-preferring constraints] up in the hierarchy. But [...] it isn't clear which of the [winner-preferring constraints] should be promoted — perhaps all of them, or perhaps just one. Other data might require one of the [winner-preferring constraints] to be dominated by one of the [loser-preferring constraints]. [The current comparative row] gives no basis for choosing.”

As a matter of fact, the issue of the convergence of the GLA or of the GLA_{min} has remained open in the literature for various years, until recently Pater (2008) has exhibited a counterexample. He reports that the GLA run on the input comparative tableau (82) starting from the null initial vector and the rows sampled in random order keeps increasing the components of its current ranking vector, without ever converging to an OT-compatible vector.¹³ Since every row of the comparative tableau

¹³ A small caveat is in order here. Pater (2008) tests the GLA on his counterexample (82) using the Praat implementation of the GLA (with standard settings: uniform sampling of the rows of the tableau; constraints initially equally ranked with

(82) contains a unique L, then the GLA and the GLA_{\min} behave in the same way on this tableau; hence, (82) also provides a counterexample against the GLA_{\min} .

$$(82) \quad \mathbf{A} = \begin{bmatrix} C_1 & C_2 & C_3 & C_4 & C_5 \\ W & L & W & & \\ & W & L & W & \\ & & W & L & W \\ & & & W & L \end{bmatrix}$$

Pater (2008) does not attempt to explain why is it that the GLA fails on this particular example (82). Here is my explanation. To get started, consider the beginning (83) of a possible run of the GLA on Pater's counterexample (82) starting from the null initial vector. Suppose that at the first iteration, the GLA gets the first row of Pater's tableau. Since the null vector is not OT-compatible with that row, update is performed. As prescribed by (78), the ranking values of the winner-preferring constraints C_1 and C_3 are promoted by 1 and the ranking value of the loser-preferring constraint C_2 is demoted by 1. Equivalently, the current ranking vector is updated by component-wise sum with the update vector corresponding to the first comparative row, namely the vector that has a 1 in correspondence of the two winner-preferring constraints C_1 and C_3 , has a -1 in correspondence of the loser-preferring constraint C_2 and has 0's elsewhere. Suppose that at the second iteration the GLA gets the second row of Pater's tableau. Since the current ranking vector is not OT-compatible with that row, update is performed. As prescribed by (78), the ranking values of the winner-preferring constraints C_2 and C_4 are promoted by 1 and the ranking value of the loser-preferring constraint C_3 is demoted by 1. Equivalently, the current ranking vector is updated by component-wise sum with the update vectors corresponding to the second comparative row, namely the vector that has a 1 in correspondence of the two winner-preferring constraints C_2 and C_4 , has a -1 in correspondence of the loser-preferring constraint C_3 and has 0's elsewhere. And so on.

$\theta_1^{\text{init}} = \dots = \theta_5^{\text{init}} = 100$; default setting of the parameters, namely $\sigma = 2.0$ and $\eta = 0.1$); see Boersma (1999). The version of the GLA implemented in Praat differs from the one I am describing here under two respects. The *first* difference is that the GLA implemented in Praat uses condition (40), whereby the current ranking vector at each iteration is perturbed with a small additive noise. Boersma suggests that this additive noise might be useful in order to model the acquisition of languages with variation; but here I am ignoring the issue of variation, and I have thus gotten rid of the additive noise. The *second* difference is that the GLA implemented in Praat uses a slightly different notion of OT-compatibility in order to decide in step 2 of the on-line algorithm (41) whether to update or not its current ranking vector. More precisely, it uses the notion of OT-compatibility from Tesar and Smolensky (2000), that allows for multiple constraints to be assigned to the same stratum with the corresponding tie resolved additively. As noted in footnote 4, this notion of OT-compatibility is different from the notion of OT-compatibility in (27) used in this work. Despite these two differences between the version of the GLA implemented in Praat and the version considered here, we can still use the Praat implementation to test the behavior of the variant of the GLA considered here. We just have to run the Praat implementation with the specifications in (i). The idea is that, since the components of the initial ranking vector are all distinct and fractional and since plasticity is integer, then the components of the current ranking vector will stay all distinct throughout learning; thus the notion of OT-compatibility built into the Praat implementation of the GLA coincides with the notion of OT-compatibility (27) used in this paper.

- (i) How to simulate the update rule (78) in Praat:
 - a. integral plasticity (e.g. $\eta = 1$);
 - b. no noise (i.e. $\sigma = 0$);
 - c. an initial vector with components $\theta_1^{\text{init}}, \dots, \theta_n^{\text{init}}$ distinct and fractional.

By using the specifications in (i), we can thus in particular test the behavior of the variant of the GLA considered in this paper on Pater's tableau (82). I have indeed run the GLA on Praat with the specifications (ia) and (ib) and the initial ranking vector (iia), that satisfies the conditions in (ic). I have got back the ranking vector (iib), that corresponds to the wrong hierarchy $C_3 \gg C_4 \gg C_1 \gg C_2 \gg C_5$. Indeed, the components of the current ranking vector just keep increasing without ever converging, just as in Pater's simulations.

- (ii) a. $\theta^{\text{init}} = (100.08, 100.06, 100.04, 100.02, 100.00)$.
- b. $\theta = (5837.684, 5837.126, 5838.516, 5838.234, 5836.696)$.

This observation also shows that the failure of the GLA on Pater's comparative tableau (82) has nothing to do with the internal randomization of the version of the algorithm defined by Boersma and tested by Pater.

$$(83) \quad \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 \\ \begin{bmatrix} W & L & W & & \\ & W & L & W & \\ & & W & L & W \\ & & & W & L \\ & & & & L \end{bmatrix} & \Rightarrow & \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} & \xrightarrow{\text{row 1}} & \begin{bmatrix} +1 \\ -1 \\ +1 \\ 0 \\ 0 \end{bmatrix} \\ & & & \xrightarrow{\text{row 2}} & \begin{bmatrix} +1 \\ -1 \\ +1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ +1 \\ -1 \\ +1 \\ 0 \end{bmatrix} \\ & & & \xrightarrow{\text{row 3}} & \begin{bmatrix} +1 \\ -1 \\ +1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ +1 \\ -1 \\ +1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ +1 \\ -1 \\ +1 \end{bmatrix} \\ & & & \xrightarrow{\text{row 4}} & \begin{bmatrix} +1 \\ -1 \\ +1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ +1 \\ -1 \\ +1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ +1 \\ -1 \\ +1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ +1 \\ -1 \end{bmatrix} \\ & & & & \rightarrow \dots \end{matrix}$$

The paper-and-pencil simulation in (83) shows that the ranking vector θ entertained by the GLA at a generic iteration must have the shape in (84), for some nonnegative coefficients $\alpha_1, \alpha_2, \alpha_3, \alpha_4$. Namely, θ must be obtained by adding together the four column vectors that appear in (84), each multiplied by a nonnegative constant α_i . The i th column vector in (84) corresponds to the i th row of Pater’s comparative tableau (82), in the sense that it is obtained by replacing W’s, L’s and E’s in that row with 1’s, -1 ’s and 0’s respectively. And the corresponding coefficient α_i in (84) represents the number of times that, up to that iteration, the GLA has updated its current ranking vector θ in response to a failure in accounting for the i th row of Pater’s comparative tableau.

$$(84) \quad \begin{matrix} \text{\# of updates in response to the} & \dots & \text{\# of updates in response to the} \\ \text{1st row of the tableau (82)} & & \text{4th row of the tableau (82)} \\ \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \end{bmatrix} = \alpha_1 \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 \\ 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{bmatrix} + \alpha_4 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} \\ \text{update vector that} & \dots & \text{update vector that} \\ \text{corresponds to the 1st} & & \text{corresponds to the 4th} \\ \text{row of the tableau (82)} & & \text{row of the tableau (82)} \end{matrix}$$

Adding up the corresponding components in (84), we conclude that the search space of the GLA run on Pater’s counterexample (82) starting from the null initial vector is a subset of the set of ranking vectors of the form (85), for nonnegative coefficients $\alpha_1, \alpha_2, \alpha_3, \alpha_4$.

$$(85) \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 - \alpha_1 \\ \alpha_1 + \alpha_3 - \alpha_2 \\ \alpha_2 + \alpha_4 - \alpha_3 \\ \alpha_3 - \alpha_4 \end{bmatrix}, \quad \alpha_1, \alpha_2, \alpha_3, \alpha_4 \geq 0$$

There is only one ranking OT-compatible with Pater's tableau (82), namely $C_1 \gg C_2 \gg C_3 \gg C_4 \gg C_5$. Thus, a ranking vector $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$ is OT-compatible with Pater's comparative tableau iff it univocally represents this ranking, namely it satisfies the four strict inequalities $\theta_1 > \theta_2 > \theta_3 > \theta_4 > \theta_5$. By virtue of the characterization (85) of the ranking vectors entertained by the GLA, these four strict inequalities can be rewritten as the four strict inequalities (86), in terms of the coefficients $\alpha_1, \alpha_2, \alpha_3, \alpha_4$.

$$(86) \begin{aligned} \theta_1 > \theta_2 &\Rightarrow \alpha_1 > \alpha_2 - \alpha_1 \\ \theta_2 > \theta_3 &\Rightarrow \alpha_2 - \alpha_1 > \alpha_1 + \alpha_3 - \alpha_2 \\ \theta_3 > \theta_4 &\Rightarrow \alpha_1 + \alpha_3 - \alpha_2 > \alpha_2 + \alpha_4 - \alpha_3 \\ \theta_4 > \theta_5 &\Rightarrow \alpha_2 + \alpha_4 - \alpha_3 > \alpha_3 - \alpha_4. \end{aligned}$$

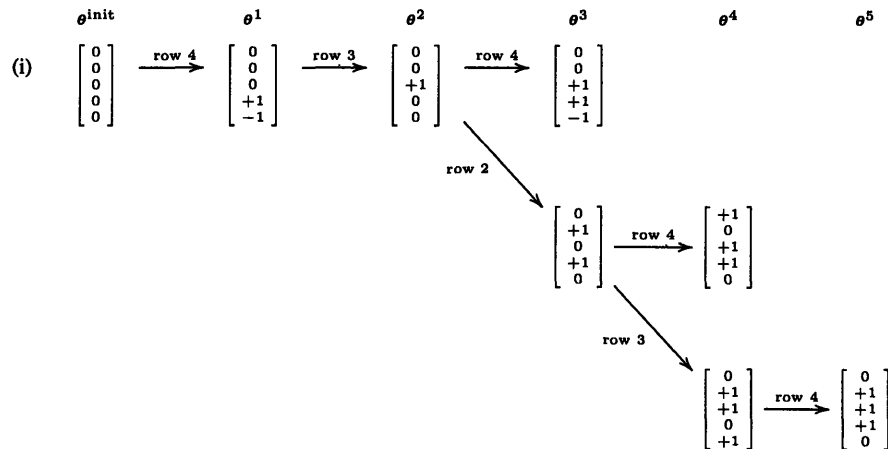
Crucially, the four strict inequalities in (86) are not feasible, namely there exist no coefficients $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ that satisfy all four of them. Here is a way to see that. If the fourth inequality in (86) is subtracted from the second one, we get (87a); if the third and the fourth inequalities in (86) are summed together, we get (87b); if the first and the fourth inequalities in (86) are summed together and then the result divided by 2, we get (87c); if the two inequalities (87b) and (87c) are summed together, we get (87d). Since (87d) contradicts (87a), then the four strict inequalities in (86) are unfeasible.

$$(87) \begin{aligned} \text{a. } &-2\alpha_1 + \alpha_2 + \alpha_3 - 2\alpha_4 > 0 \\ \text{b. } &\alpha_1 - \alpha_2 + \alpha_4 > 0 \\ \text{c. } &\alpha_1 - \alpha_3 + \alpha_4 > 0 \\ \text{d. } &-2\alpha_1 + \alpha_2 + \alpha_3 - 2\alpha_4 < 0. \end{aligned}$$

In conclusion, the reason why the GLA fails on Pater's comparative tableau (82) is as follows: the search space of the GLA is limited to ranking vectors of the form (85); but no such ranking vector is OT-compatible with Pater's comparative tableau, namely satisfies the inequalities (86). In other words, the GLA fails because it struggles to reach a ranking vector that lies behind its reach.^{14,15}

¹⁴The preceding discussion explains why the GLA fails on Pater's counterexample (82), but it does not explain why in the case of this counterexample the components of the current ranking vector entertained by the GLA just keep increasing. An explanation of this fact follows from two general properties of promotion-demotion update rules studied in chapter 5. First, the components of the current ranking vector cannot decrease below a certain value, namely they are lower bounded; see claim 12. Second, the algorithm cannot entertain the same ranking vector twice within the same run; see claim 14.

¹⁵As noted in footnote 4, Tesar and Smolensky (2000) introduce a notion of OT-compatibility alternative to the standard notion of OT-compatibility (27) considered in this work. According to their alternative notion of OT-compatibility, ties among equally ranked constraints are resolved additively. Let GLA' be the variant of the GLA obtained by using this alternative notion of OT-compatibility in step 2 of the on-line algorithm (41). As noted in footnote 13, Praet actually implements the GLA' , not the GLA. Contrary to the GLA, the GLA' works fine on Pater's counterexample (82): as shown by the diagram in (i), the GLA' converges after at most 5 updates, no matter the order in which the rows of Pater's comparative tableau (82) are fed to the algorithm.



4.2.6.2 On the analysis of promotion-demotion update rules

It is useful to restate this discussion of Pater’s counterexample in a slightly more abstract way, that paves the way to the developments of chapter 5. A run of the on-line algorithm (41) with any update rule on an input comparative tableau $A \in \{L, E, W\}^{m \times n}$ describes a path of ranking vectors as in (88a), where θ^{init} is the initial ranking vector, θ^{fin} is the final ranking vector of that run (assuming that the algorithm converges in that run) and $\theta^t = (\theta_1^t, \dots, \theta_n^t)$ is the current ranking vector entertained by the algorithm at a generic time t . There are two ways of keeping track of “time”, and thus of interpreting this superscript t . One way is to interpret time as counting *all* the loops (41), also those where the comparative row fed to the algorithm is OT-compatible with the current ranking vector and thus no update is performed. Another way is to interpret time as counting *only* those loops (41) that trigger an update. Let me adopt the latter strategy. Thus, θ^t is the ranking vector entertained by the algorithm after t updates; and the identity $\theta^T = \theta^{\text{fin}}$ means that it took T updates in that run for the algorithm to reach a final ranking vector θ^{fin} OT-compatible with the input tableau. Let α_i^t be the number of updates out of the first t updates that were triggered by a failure in accounting for the i th row of the input comparative tableau A , for every $i = 1, \dots, m$. Let me collect the m numbers $\alpha_1^t, \dots, \alpha_m^t$ into an m -tuple $\alpha^t = (\alpha_1^t, \dots, \alpha_m^t)$. I will call α^t the DUAL VECTOR at time t and I will call its i th component α_i^t the i th DUAL VARIABLE at time t . I will thus also call $\theta^t = (\theta_1^t, \dots, \theta_n^t)$ the PRIMAL VECTOR at time t and I will call its h th component θ_h^t the h th PRIMAL VARIABLE at time t . A run of the on-line algorithm (41) thus describes, besides the PRIMAL PATH (88a), also the DUAL PATH (88b).¹⁶

The success of the GLA’ on Pater’s comparative tableau fits well with the account given above for the failure of the GLA on that same tableau. In fact, the moment we allow the algorithm to exploit ties by switching from the standard notion of OT-compatibility to the alternative notion of OT-compatibility considered in Tesar and Smolensky (2000), we effectively enlarge the set of rankings that account for Pater’s tableau to also include non-total rankings like the ones in (ii). The success of the GLA’ thus just says that these three rankings can indeed be represented by ranking vectors that belong to the search space (85) of the algorithm.

- (ii) a. $\{C_3, C_4\} \gg \{C_1, C_2\} \gg \{C_5\}$.
- b. $\{C_1, C_3, C_4\} \gg \{C_2, C_5\}$.
- c. $\{C_2, C_3, C_4\} \gg \{C_1, C_5\}$.

Thus, Pater’s comparative tableau (82) provides a counterexample against the GLA (with the standard notion of OT-compatibility) but does not provide a counterexample against the GLA’ (with the alternative notion of OT-compatibility considered by Tesar and Smolensky). Yet, consider the comparative tableau in (iiia). Suppose that the rows of this comparative tableau are fed to the GLA’ in the fixed order row 3 → row 2 → row 1. The first two passes through the data are given in (iiib). Note that at the end of the second pass, the GLA’ is entertaining a ranking vector that represents the same non-total ranking represented by the initial null vector. Thus, the GLA’ doesn’t converge.

(iii) a.
$$\begin{bmatrix} W & L & W & W \\ & W & L & W \\ & & W & L \end{bmatrix}$$

b.
$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \xrightarrow{\text{row 3}} \begin{bmatrix} 0 \\ 0 \\ +1 \\ -1 \end{bmatrix} \xrightarrow{\text{row 2}} \begin{bmatrix} 0 \\ +1 \\ 0 \\ 0 \end{bmatrix} \xrightarrow{\text{row 1}} \begin{bmatrix} +1 \\ 0 \\ +1 \\ +1 \end{bmatrix} \xrightarrow{\text{row 3}} \begin{bmatrix} +1 \\ 0 \\ +2 \\ 0 \end{bmatrix} \xrightarrow{\text{row 2}} \begin{bmatrix} +1 \\ +1 \\ +1 \\ +1 \end{bmatrix} \xrightarrow{\text{row 1}} \begin{bmatrix} +1 \\ +1 \\ +1 \\ +1 \end{bmatrix}$$

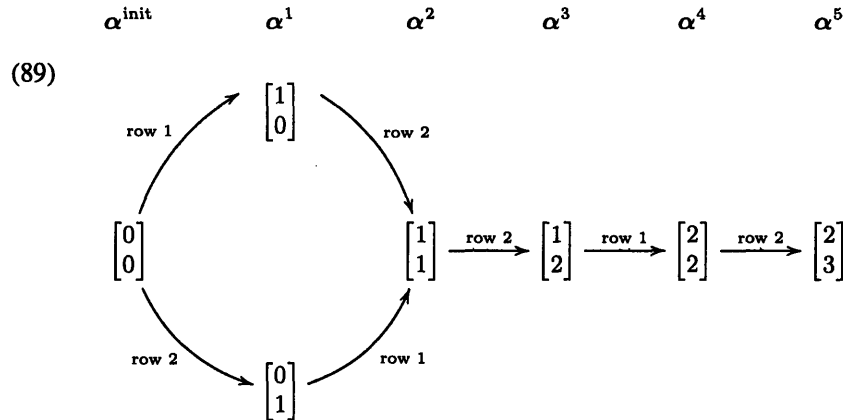
c.
$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \xrightarrow{\text{row 1}} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \xrightarrow{\text{row 2}} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \xrightarrow{\text{row 3}} \begin{bmatrix} 0 \\ 0 \\ +1 \\ -1 \end{bmatrix} \xrightarrow{\text{row 1}} \begin{bmatrix} 0 \\ 0 \\ +1 \\ -1 \end{bmatrix} \xrightarrow{\text{row 2}} \begin{bmatrix} 0 \\ +1 \\ 0 \\ 0 \end{bmatrix} \xrightarrow{\text{row 3}} \begin{bmatrix} 0 \\ +1 \\ +1 \\ -1 \end{bmatrix}$$

It is interesting to note that the GLA’ does converge if the rows of the comparative tableau (iiia) are fed in a different order. For instance, the GLA’ does converge after two passes through the data in the rows of the comparative tableau (iiia) are fed to the algorithm in the fixed order row 1 → row 2 → row 3, as shown in (iiic).

¹⁶The terminology PRIMAL and DUAL comes from the connection between on-line algorithms for linear classification and duality theory for linear programming.

$$(88) \quad \begin{array}{l} \text{a.} \quad \theta^{\text{init}} \longrightarrow \theta^1 \longrightarrow \theta^2 \longrightarrow \dots \longrightarrow \theta^t \longrightarrow \dots \longrightarrow \theta^T = \theta^{\text{fin}} \\ \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ \text{b.} \quad 0 \longrightarrow \alpha^1 \longrightarrow \alpha^2 \longrightarrow \dots \longrightarrow \alpha^t \longrightarrow \dots \longrightarrow \alpha^T \end{array}$$

To illustrate, I provide in (89) the dual paths corresponding to the primal paths in (80). The dual vectors have two components, since the input comparative tableau (15) has $m = 2$ rows. The initial dual vector is of course the null vector, since at the beginning no row has triggered a single update. The dual variables at any time are of course non-negative integers. And the dynamics over time of the dual variables is very simple, since each of them can only increase over time.



If a given update rule yields a simple dynamics of the primal vectors θ^t , then the behavior of the corresponding on-line algorithm (41) is best characterized in terms of its primal paths. This is indeed the case for demotion-only update rules. These rules yield a very simple monotonic dynamics of the primal variables, since they can only decrease over time. For this reason, the corresponding on-line algorithm can be studied in terms of primal paths, as done indeed throughout sections 4.2.2-4.2.5. The situation with promotion-demotion update rules is very different. The corresponding dynamics of the primal vectors is much more complicated because it is non-monotonic: as illustrated for example in (80), a given primal variable can oscillate up and down over time. The behavior of the on-line algorithm (41) corresponding to promotion-demotion update rules is thus better characterized in terms of its dual paths. In fact, the dual dynamics is always monotonic, since dual variables can only increase over time. Furthermore, the dual dynamics completely determines the primal dynamics. In the rest of this subsection, I comment on the latter fact in some detail.

- (90) a. Demotion-only update rules: analysis at the primal level (88a);
- b. promotion-demotion update rules: analysis at the dual level (88b).

To start, note that the identity (91) trivially holds: the number of updates t is of course given by the sum of the number of updates α_1^t triggered by the first row plus the number of updates α_2^t triggered by the second row and so on. This identity entails in particular that the total number of updates T in a given run can be described as the sum of the components of the final dual vector α^T corresponding to that run.

$$(91) \quad t = \sum_{i=1}^m \alpha_i^t$$

Furthermore, the dual vector α^t at time t fully determines the primal vector θ^t at time t . Let me illustrate this point in detail, starting from the specific case of Boersma's update rule (78) and slowly moving toward the general case. This update rule fits into the general scheme $\theta^{\text{new}} = \theta^{\text{old}} + \bar{a}$ in (46), with the update vector $\bar{a} = (\bar{a}_1, \dots, \bar{a}_n)$ corresponding to a comparative row a that triggers that update is equal to W (or to L or to E , respectively). Thus, in this case the update vector only

depends on the current comparative row, and we can speak of the update vector $\bar{a} \in \{-1, 0, +1\}^n$ CORRESPONDING to a comparative row $a \in \{L, E, W\}^n$. Let \bar{A} be the numerical matrix obtained by organizing one underneath the other all the update vectors \bar{a} corresponding to all the rows a of the input tableau. For instance, Pater's comparative tableau A in (82) is paired up with the numerical matrix \bar{A} in (92), obtained by replacing every W with $+1$, every L with -1 and every E with 0 (for readability, I omit E 's in A and 0 's in \bar{A}).

$$(92) \quad A = \begin{bmatrix} W & L & W & & \\ & W & L & W & \\ & & W & L & W \\ & & & W & L \end{bmatrix} \implies \bar{A} = \begin{bmatrix} +1 & -1 & +1 & & \\ & +1 & -1 & +1 & \\ & & +1 & -1 & +1 \\ & & & +1 & -1 \end{bmatrix}$$

The four rows of the matrix \bar{A} are by construction the four update vectors that appear in (84). In other words, the identity (84) can be rewritten as in (93): the primal vector θ^t is a linear combination¹⁷ of the rows of the matrix \bar{A} with coefficients provided by the components of the corresponding dual vector α^t .¹⁸

$$(93) \quad \theta^t = \alpha_1^t \begin{bmatrix} \text{1st row of } \bar{A} \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} + \dots + \alpha_m^t \begin{bmatrix} \vdots \\ \vdots \\ \vdots \\ \text{mth row of } \bar{A} \end{bmatrix}$$

It is useful to rewrite (93) component-wise. For ease of presentation, let me preliminarily introduce the piece of notation in (94): given two vectors $\mathbf{v} = (v_1, \dots, v_d)$, $\mathbf{w} = (w_1, \dots, w_d) \in \mathbb{R}^d$ both with the same number d of components, their SCALAR PRODUCT $\langle \mathbf{v}, \mathbf{w} \rangle$ is the sum of the components of one of the two vectors weighted by the corresponding components of the other.

$$(94) \quad \langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + \dots + v_d w_d = \sum_{h=1}^d v_h w_h$$

Let $\bar{A}_1, \dots, \bar{A}_n \in \mathbb{R}^m$ be the n columns of the matrix \bar{A} obtained as above by ordering all the update vectors one underneath the other. The identity (93) can be read component-wise as follows: the first component θ_1^t of the current ranking vector is the sum of the first component of the first row of \bar{A} multiplied by α_1^t plus the first component of the second row of \bar{A} multiplied by α_2^t plus the first component of the third row of \bar{A} multiplied by α_3^t , and so on. Equivalently, the first component θ_1^t of the current ranking vector is the scalar product between the first column \bar{A}_1 with the dual vector α^t . In the general case, we get the identity (95) for every $k = 1, \dots, n$.

$$(95) \quad \theta_k^t = \langle \bar{A}_k, \alpha^t \rangle$$

The identities (93) and (95) extend to any update rule that fits into the general scheme $\theta^{\text{new}} = \theta^{\text{old}} + \bar{a}$ in (46). If the update vector \bar{a} only depends on the current comparative row and not on the current ranking vector, then each comparative row of the input tableau is paired up with a single update vector, and the analysis just presented can be repeated verbatim. Otherwise, assumption (43b) ensures that each comparative row is paired up only with a finite number of update vectors. In this case, the preceding analysis can be straightforwardly readapted, by introducing a dual variable for each update vector. By (45), a theory of an update rule for the on-line algorithm (41) should provide a bound on the worst case number of updates T and should characterize the final primal vector θ^T . By the two identities (91) and (93)/(95), this task can be carried out by investigating properties of the dual dynamics of the corresponding on-line algorithm. Dual variables will indeed turn out to be very useful for the developments of chapter 5.

¹⁷Actually, a *conic* combination, since the coefficients $\alpha_1^t, \dots, \alpha_m^t$ are nonnegative.

¹⁸With standard linear Algebra notation, the identity (93) says that $\theta^t = \bar{A}^T \alpha^t$, namely that the primal vector θ^t at time t is the row-by-column product between the transpose of the matrix \bar{A} , obtained by organizing the update vectors one underneath the other, and the corresponding dual vector α^t .

4.3 The OT on-line algorithm: a modeling perspective

The preceding section has introduced the OT on-line algorithm from a purely computational perspective. This section looks at the algorithm from a modeling perspective. In particular, subsection 4.3.4 makes the point that none of the update rules reviewed from the literature in the preceding section is suitable in order to model the early stage of the acquisition of phonology prior to morphological awareness, as described in Hayes (2004).

4.3.1 Initialization

A crucial ingredient of the OT on-line model (32) is the initialization of the current ranking to a specific initial ranking \gg^{init} . There is wide agreement in the literature that the initial ranking should rank markedness constraints above faithfulness constraints, as stated in (96).¹⁹ For instance, Fikkert and De Hoop (2009, p. 325) write: “The recurrent pattern in child language data is that children’s output is considerably less marked than the corresponding adult target forms. This is true both for segmental, syllabic and higher prosodic structure. Hence, the starting hypothesis in much research on phonological acquisition is that children begin with with markedness constraints outranking faithfulness constraints.” See Smolensky (1996a,b) for theoretical arguments in favor of (96); see Jusczyk et al. (2002) for empirical evidence; see Davidson et al. (2004) for a review.

$$(96) \quad \mathcal{M} \gg^{\text{init}} \mathcal{F}$$

Assumption (96) has been refined in various ways: Smith (2000) and Revithiadou and Tzakosta (2004) argue that positional faithfulness constraints should be initially ranked on top of the corresponding general faithfulness constraints; and McCarthy (1998) argues that (96) only applies to input-output faithfulness constraints, while output-output faithfulness constraints should start out on top.

4.3.2 Choice of the underlying form

Hayes (2004) reviews the relevant psycholinguistic literature and concludes that the knowledge of an eight-to-ten month old child can be characterized by means of the two properties (97). He dubs this developmental stage the EARLY STAGE of the acquisition of phonology.

- (97) a. KNOWLEDGE OF PHONOTACTICS. “At more or less [eight to ten months], infants start to acquire knowledge of the legal [...] sequences of their language. [...] In carefully monitored experimental situations, eight-to-ten month old infants come to react differently to legal phoneme sequences in their native language than to illegal or near-illegal ones”.
- b. LACK OF KNOWLEDGE OF ALTERNATIONS. “Certainly we can say that there are at least some morphological processes which are acquired long after the system of contrasts and phonotactics is firmly in place, and it seems a reasonable guess that in general, the learning of patterns of alternation lags the learning of the contrast and phonotactic systems”.

That morphological awareness does indeed lag behind knowledge of phonotactics, as stated in (97), is shown by a number of case studies. Here is one of these case studies, taken from Kazazis (1969) via Jesney and Tessier (2007).²⁰ Adult Greek phonotactics displays the restriction in (98), both in non-derived and derived environments.

- (98) a. Only palatals occur before front vowels (e, i):
- | | |
|--------------|--------------------|
| /ce/ → [ce], | /ec+ete/ → [ecete] |
| /xe/ → [ce], | /ex+ete/ → [ecete] |

¹⁹The only exception I know of is Hale and Reiss (1998).

²⁰This case is also mentioned in Hayes (2004), who notes that an analogous case is discussed in Bernhardt and Stemberger (1998, p. 641).

- b. Only velars occur before non-front vowels (V):
 /cV/ → [xV], /ec+Vte/ → [ex+Vte]
 /xV/ → [xV], /ex+Vte/ → [ex+Vte]

Kazazis (1969) documents the U-shaped learning path in (99). At an initial stage (99a), the child has acquired the phonotactics in (98) and the fact that it applies both to derived and non-derived environments. At a later stage (99b), the child un-learns the correct phonotactics in derived environments. Only at a later stage (99c) does the child recover from the mistake and goes back to the correct phonotactics.

- (99) a. Stage 1: everything is correct.
 b. Stage 2: everything is correct in underived forms;
 but mistakes in derived forms, namely /ex+ete/ → [exete].
 c. Stage 3: everything is correct again.

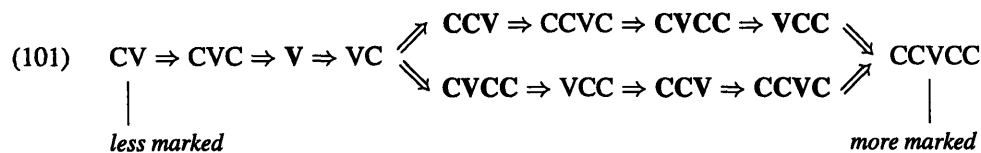
This learning path (99) is immediately accounted for if we assume that morphological awareness kicks in at a later stage. Assume the four constraints *C, *XE, IDENTIO and IDENTOO. Assume that at stage 1 morphological awareness has not started yet. Thus, there is no difference between underived and derived forms and the faithfulness constraint IDENTOO does not play any role. This stage corresponds to the ranking (100a). At stage 2, morphological awareness kicks in and IDENTOO thus becomes active. If we assume that it starts out top ranked, then this second stage corresponds to the ranking (100b). Since this ranking is incorrect, learning is triggered, that leads to the correct ranking (100c) corresponding to stage 3.

- (100) a. Stage 1: *XE ≫ *C ≫ IDENTIO
 b. Stage 2: IDENTOO ≫ *XE ≫ *C ≫ IDENTIO
 c. Stage 3: *XE ≫ IDENTOO ≫ *C ≫ IDENTIO

A child at this early stage when he knows no morphology cannot take advantage of alternations. Thus, he has no information about underlying forms. Hence, the best he can do at this early stage is to posit fully faithful underlying forms. In conclusion, assumption (35) makes sense not only from the computational perspective but also from the modeling perspective, at least as long as we limit ourselves to the task of modeling the early stage of the acquisition of phonology, when morphological awareness is lagging behind.

4.3.3 A toy example

In this section, I illustrate the on-line model of the acquisition of phonology by reviewing a toy example from Boersma and Levelt (2000); Curtin and Zuraw (2002) provide another almost identical example. Levelt et al. (2000) investigate the production of stressed syllables by twelve children acquiring Dutch. Adult Dutch phonotactics allows all nine syllable types. They report the remarkable finding that, among the many learning paths in principle possible, only the two specific paths in (101) are attested. Each syllable type in the diagram represents an intermediate stage in the acquisition path at which the child produces that syllable type together with all the syllable types at its left but none of the syllable types at its right. For example, the entry V in (101) represents a stage at which the child produces CV, CVC and V, but neutralizes any other syllable type. The two acquisition paths in (101) can roughly be described as slowly moving from the “least marked” to the “most marked” syllable type.

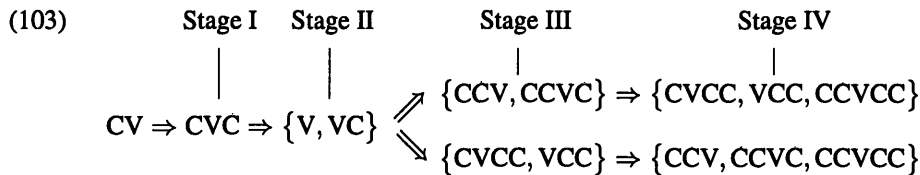


The relevant portion of the universal specifications is provided in (102). The set of underlying forms \mathcal{X} and the set of surface forms \mathcal{Y} coincide and consist of the nine syllable types. The set of candidates

$Gen(x)$ for any syllable type x is obtained by arbitrary consonant deletion and epenthesis in the two margins of the syllable. The constraint set contains the four standard markedness constraints for syllable types plus a unique faithfulness constraint (defined as the pointwise sum of traditional DEP and MAX).

- (102) a. $\mathcal{X} = \mathcal{Y} = \{CV, CVC, V, VC, CCV, CCVC, CVCC, VCC, CCVCC\}$
 b. $Gen(x) = \mathcal{Y}$, for every syllable type $x \in \mathcal{X}$
 c. NOCODA: assigns a violation for every syllable that is closed;
 ONSET: assigns a violation for every syllable that lacks an onset;
 COMPCODA: assigns a violation for every syllable that has a complex coda;
 COMPONSET: assigns a violation for every syllable that has a complex onset;
 FAITH: assigns a violation for every input segment that does not have a correspondent in the output or vice versa.

It is well known that the boldfaced intermediate stages in (101) do not correspond to any whole language in the typology defined by the universal specifications (102); see Levelt et al. (2000) and Albright et al. (2007) for further discussion. For this reason, Boersma and Levelt (2000) decide to ignore these problematic intermediate stages and limit themselves to the simplified learning path in (103), with only four stages (besides the initial one).



Weijer (1997) provides the frequencies (104) of the nine syllable types in a child-directed corpus of 112,926 primary stressed Dutch syllables.

(104)	CV:	44.81%	V:	3.85%	CCV:	1.38%
	CVC:	32.05%	CVCC:	3.25%	VCC:	0.42%
	VC:	11.99%	CCVC:	1.98%	CCVCC:	0.26%

From (104), the frequencies of violations of the four markedness constraints in (102c) are immediately computed as in (105), by summing up the frequencies of all the forms that violate a given constraint. The frequencies of violations in (105) of course reflect both the lexical frequencies in (104) as well as the entailments built into the definition of the constraints (e.g. every time COMPCODA is violated, NOCODA is violated too).

(105)	NOCODA:	44.95%	COMPCODA:	3.93%
	ONSET:	16.26%	COMPONSET:	3.62%

Each one of the four intermediate stages in the simplified paths in (103) corresponds to one of the four markedness constraints, in the sense that that stage arises as soon as the corresponding markedness constraint drops below the faithfulness constraints. I illustrate this correspondence in (106) for the upper path in (103).

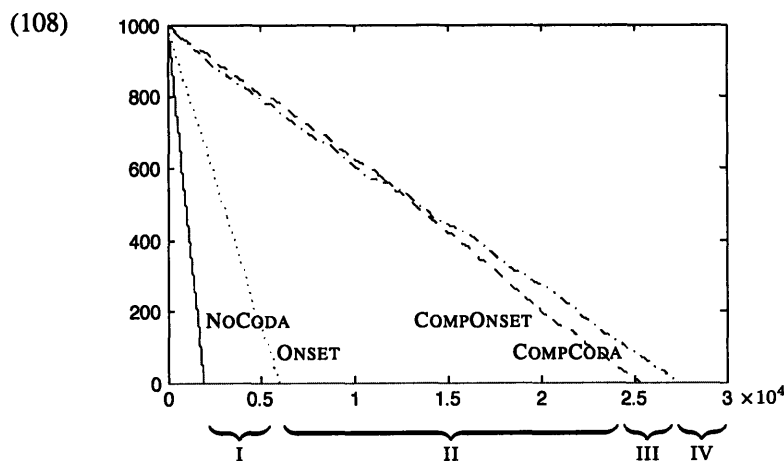
(106)	Stage I	⇔	FAITH	≫	NOCODA
	Stage II	⇔	FAITH	≫	ONSET
	Stage III	⇔	FAITH	≫	COMPONSET
	Stage IV	⇔	FAITH	≫	COMPCODA

As noted in Levelt and van de Vijver (1998), once the input frequencies (104) are organized by constraint as in (105), the order of acquisition in (103) is immediately accounted for through the correspondence in (106): we expect syllables with simple codas to appear first, because the corresponding constraint NOCODA is the one violated most often and thus the fastest one to drop below FAITH; we expect onsetless syllables to appear next, because the corresponding constraint ONSET is the next most often violated constraint and thus the next fastest one to drop below FAITH; we

expect complex margins to appear thereafter, once the corresponding constraints *COMP*CODA and *COMP*ONSET drop below *FAITH*. The on-line model (33) straightforwardly captures this intuition with pretty much any update rule. For concreteness, consider the specific implementation of the model described in (107).

- (107) a. *Initial ranking vector*: $\theta_{\text{FAITH}}^{\text{init}} = 0$
 $\theta_{\text{CODA}}^{\text{init}} = \theta_{\text{ONSET}}^{\text{init}} = \theta_{\text{CCODA}}^{\text{init}} = \theta_{\text{CONSET}}^{\text{init}} = 1000$
- b. *Selection of the underlying form in step 1*: according to (34).
- c. *Selection of the loser candidate in step 1*: according to (39).²¹
- d. Update rule used in step 3: if the current ranking vector θ is not OT-compatible with the current underlying/winner/loser form triplet (x, y^*, y) , demote by 1 the ranking value of all loser-preferring markedness constraints.

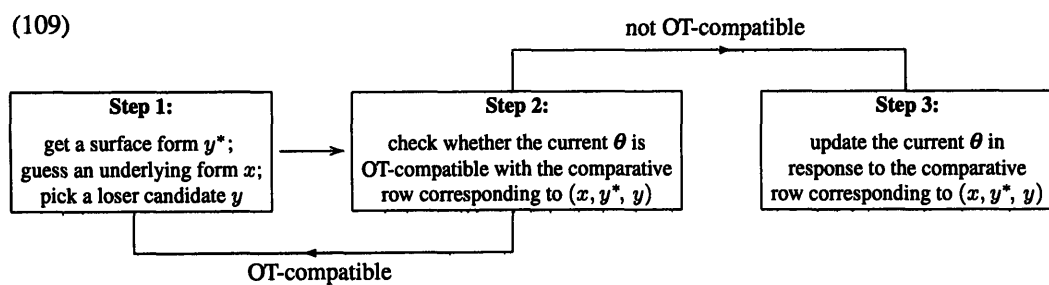
The diagram in (108) reports the dynamics of the ranking values of the five constraints over time in one run of the model. The faithfulness constraint stays put at its null initial ranking value. The markedness constraints fall down with a slope that is determined by the violation frequencies in (105). The succession of the four stages (103) is thus straightforwardly predicted.



I'll consider quite more complicated test cases in chapter 6, once I'll have refined the description of the model in chapter 5.

4.3.4 How to choose update rules

To complete the description of the on-line model (33) for the acquisition of phonology, repeated in (109), we need to go back to the small arsenal of update rules (47) reviewed in the preceding section, and decide which one of them is best suited to be used in step 3. Let me consider each one of the update rules listed in (47) in turn.



Boersma (1997) notes that non-gradual update rules, such as (66), are not suitable update rules for

²¹More precisely, I have broken ties according to the order with which the constraints are listed in (102). Namely, if say *NOCODA* and *ONSET* have the same current ranking value, the corresponding ranking considered in (39) ranks *NOCODA* above *ONSET* because the former is listed before the latter in (102).

the on-line model (109) of the acquisition of phonology. The idea is as follows. All the action in the on-line model is carried out by the choice of the initial ranking vector θ^{init} . Thus, we want update rules that are very sensitive to the properties of the initial ranking vectors. Non-gradual update rules are less sensitive to the properties of the initial vector than gradual update rules. This is shown for example by a comparison between the bounds on the worst-case number of updates required by the two rules: the bound for the gradual update rule (51) depends on the properties of the initial vector θ^{init} , summarized in the quantity $\Delta(\theta^{\text{init}})$ defined in (63); the bound for the non-gradual update rule (66) does not depend on the properties of the initial vector, that indeed does not figure at all in the bound.

(110) The *non-gradual* update rule (66) is not suitable for the on-line model (109) of the acquisition of phonology.

Let me suggest next that non-minimal update rules, such as (71) and (78), are not suitable update rules either for the on-line model (109) of the acquisition of phonology. The issue here is not that they are slower than minimal update rules, as noted in subsection 4.2.5. Indeed, the new update rules that I will advocate in chapter 5 will turn out to be rather slow too in the worst case. The issue is deeper than that, namely that non-minimal update rules are slower because they are “wrong”, in the sense that they do not match the intrinsic logic of OT. This intrinsic logic says that a comparative row with multiple entries equal to L is equivalent to multiple rows each with a single entry equal to L, as stated in claim 2. A proper update rule should reflect this fact, namely should have the property that a single update by a row with multiple L’s should be equivalent to the series of updates by the corresponding rows with a single L. Minimal update rules ensure this equivalence, and thus capture the intrinsic ranking logic of OT. Non-minimal update rules do not ensure this equivalence, and thus fail to capture a crucial property of the ranking logic of OT. Because of the fact that non-minimal update rules do not capture the intrinsic logic of OT, I postulate (111).

(111) The *non-minimal* update rules (71) and (78) are not suitable for the on-line model (109) of the acquisition of phonology.

My assumption (111) rules out Boersma’s non-minimal promotion-demotion update rule (78). What about the corresponding minimal update rule (79)? I would like to suggest that that update rule too does not qualify as a suitable update rule for the on-line model of the acquisition of phonology, for the same reason that underlies assumption (111) just discussed. The issue here is not that this update rule does not converge in the general case, as shown by Pater’s counterexample (82). Indeed, it might well be the case that the somewhat contrived input tableau devised by Pater never arises in a realistic setting. The issue is deeper than that, namely that this update rule fails in certain well-constructed cases because it is “wrong”, again in the sense that it does not match the intrinsic logic of OT. I will elaborate on this point at the beginning of section 5.1, where I will construct a replacement of this update rule, based on the specific properties of the ranking logic of OT. Because of the fact that the update rule (79) does not correspond to the intrinsic logic of OT, I postulate (112).

(112) The *non-convergent* update rule (79) is not suitable for the on-line model (109) of the acquisition of phonology.

Finally, let me argue that the on-line model (109) of the acquisition of phonology has no chances of working if we adopt a demotion-only update rule. By assumption (34), repeated in (113), the model assumes fully faithful underlying forms. As noted by Tesar (2008), assumption (113) makes sense from the *computational* perspective, because it cannot ever lead to a mistake (under mild assumptions on the constraint set). As noted by Hayes (2004), assumption (113) makes sense also from the *modeling* perspective, at least as long as we restrict ourselves to the task of modeling the so called EARLY STAGE of the acquisition of phonology, characterized by (some) knowledge of phonotactics but no knowledge of morphology. In fact, lack of knowledge of morphology prevents the learner at this stage from taking advantage of alternations, and thus from being able to posit any underlying form different from the fully faithful one.

(113) In step 1 of algorithm (109), given a surface form y^* , let the corresponding underlying form x be defined by $x = y^*$.

If the underlying forms are fully faithful as by (113), then the faithfulness constraints are never loser-preferring. A demotion-only update rule will therefore never modify the initial ranking values of the faithfulness constraints, since a demotion-only update rule only demotes loser-preferring constraints. In other words, whatever the phonotactics of the target language the learner has been exposed to, the final ranking values of the faithfulness constraints at the end of the early stage will be identical to the initial ranking values. This cannot be right, because phonotactics corresponding to different languages require different rankings of the faithfulness constraints. One way to make this argument more concrete is as follows. For the sake of the argument, consider a typology that has only general input-output faithfulness constraints, thus ignoring various distinctions within the set of faithfulness constraints, such as the distinction between general and positional faithfulness constraints and the distinction between input-output and output-output faithfulness constraints. In this case, it makes sense to assume that the faithfulness constraints start out equally ranked. Since the faithfulness constraints are never loser-preferring and since a demotion-only update rule only demotes loser-preferring constraints, then the faithfulness constraints will remain equally ranked. In other words, the search space of the on-line algorithm (109), implemented with assumption (113) and a demotion-only update rule, is effectively limited to ranking vectors that assign all faithfulness constraints to the same stratum, as stated in (114).

$$(114) \text{ search space } \subseteq \left\{ \theta = \begin{pmatrix} F_1 & \dots & F_\ell & M_{\ell+1} & \dots & M_n \\ \theta_1 & \dots & \theta_\ell & \theta_{\ell+1} & \dots & \theta_n \end{pmatrix} \mid \theta_1 = \dots = \theta_\ell \right\}$$

This artificial restriction of the search space imposed by the choice of the update rule might be fatal. First, there is no way for the model to account for intermediate stages where a markedness constraint is ranked in between two faithfulness constraints, despite the fact that such intermediate stages are indeed attested, as documented for example in Gnanadesikan (2004). Second, there is no way to model the sequence of two intermediate stages where two different repair strategies arise by the different relative ranking of two faithfulness constraints, despite the fact that these learning paths are indeed attested, as documented for example in Bernhardt and Stemberger (1998). Third, there is no way for the model to converge on the target language, if that language requires a markedness constraint to be ranked in between two faithfulness constraints. For these reasons, I postulate (115).

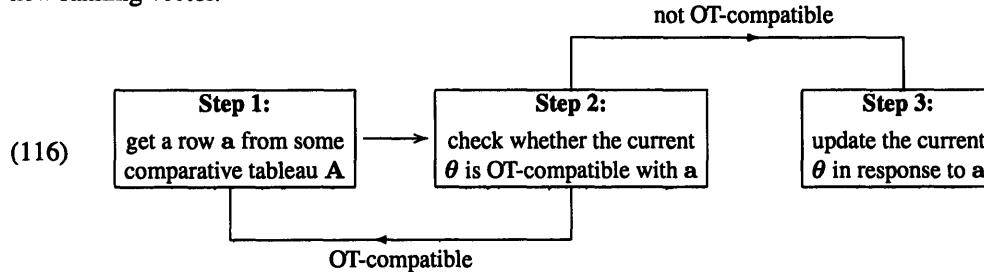
(115) The *demotion-only* update rule (51) is not suitable for the on-line model (109) of the acquisition of phonology.

Putting together (110)-(115), I conclude that none of the update rules (47) considered so far in the literature can be used to implement the on-line model of the acquisition of phonology (109). The preceding discussion furthermore suggests that one such suitable update rule should be gradual, minimal, promotion-demotion and satisfy the crucial convergence condition (44). The first crucial open issue in the theory of the OT on-line model of the acquisition of phonology is whether such update rules exist. If some such update rules can be devised, then the OT on-line model has got a chance. If no such rules can be devised, as suggested by the passage quoted in (81) from Tesar and Smolensky, then the OT on-line model needs to be abandoned and alternative models need to be pursued. I will take on this issue in the next chapter, and show that the desired update rules can indeed be constructed.

Chapter 5

How to get constraint promotion

The preceding chapter has introduced the OT on-line algorithm (41), repeated in (116): at each time, the algorithm is fed a current comparative row a ; it checks whether the current ranking vector θ is OT-compatible with that comparative row; if it is not, it updates the current ranking vector θ to a new ranking vector.



The core ingredient of the on-line algorithm (116) are *proper* update rules for step 3. Namely, update rules that provably satisfy condition (44), repeated in (117). The intuition being that condition (117) ensures that the corresponding update rule is “well tuned” with the ranking logic of OT.

(117) If the comparative rows in step 1 are sampled from an arbitrary OT-compatible input tableau A , then the algorithm can perform only a finite number of updates in step 3.

As noted in section 4.2, the only update rules currently existing in the literature that satisfy condition (117) are those that perform demotion-only. Demotion-only update rules are characterized by the two crucial properties (118). These two properties ensure that demotion-only update rules are easy to study by induction on the constraints, as in the proofs of claims 4, 5 and 7 in the preceding chapter. In fact, property (118a) ensures the existence of a ranking value that is never modified through learning (because every OT-compatible tableau contains at least one constraint that is never loser-preferrer), and that can thus be used as the base of the induction. Property (118b) ensures a monotonic dynamics of the ranking values, that yields a simple inductive step. The situation is very different for update rules that perform both promotion and demotion. The only such rules considered in the literature are Boersma’s (1997) update rules (78) and (79), that were shown by Pater’s (2008) counterexample not to satisfy condition (117). Furthermore, the simple line of analysis that works for demotion-only update rules does not seem to immediately extend to promotion-demotion update rules. Finally, Tesar and Smolensky explicitly warn against constraint promotion, in the celebrated passage quoted in (81).

- (118) a. Only the ranking value of loser-preferrer constraints is modified;
b. whenever a ranking value is modified, it is decreased.

On the other hand, subsection 4.3.4 made the point that demotion-only update rules do not seem suitable to model language acquisition, or at least that early stage of language acquisition where the learner entertains fully faithful underlying forms because of lack of knowledge of morphology.

In fact, if the underlying forms are fully faithful, then the faithfulness constraints are never loser-preferrer and thus their ranking values are never modified by a demotion-only update rule. This means in turn that, if two faithfulness constraints F, F' start out equally ranked, they will remain equally ranked, with no chance for a markedness constraint to be ranked in between. In other words, there is no way for the model to describe a learning path with the initial and final states in (119).

$$(119) \quad M \gg \{F, F'\} \longrightarrow F \gg M \gg F'.$$

Fikkert and De Hoop (2009, pp. 314-315) note that “the field of the acquisition of phonology in OT is in fact split into two subdivisions. In one division, research is based on empirical data: child language acquisition data are studied and developmental patterns are unraveled. The other division investigates learnability issues. [...] Ideally, the model mimics real language acquisition, but this is [not] of much concern in actual learnability studies. Often the term ‘learning’ is used in the latter context, while ‘acquisition’ refers to child language development [...]. So far, learnability studies have not taken actual acquisition patterns, or real learners, into account.” The situation just described illustrates well this divide between computational simplicity and modeling complexity with a case where a computational virtue turns into a modeling drawback: the computational virtue of demotion-only update rules is that they yield a very simple primal dynamics, that is easy to analyze; the modeling drawback is that this dynamics is too simple to model the observed complexity of the acquisitional path. This chapter addresses this impasse by deriving various minimal, gradual promotion-demotion update rules for the OT on-line algorithm (116) that provably satisfy condition (117), namely only trigger a finite number of updates. For the ease of exposition, I will start out by considering a specific promotion-demotion update rule, constructed in section 5.1 based on heuristic considerations on the ranking logic of OT. I will then present two different proofs that condition (117) holds for this update rule. In section 5.2, I will present a combinatoric proof, that is close in spirit to Tesar and Smolensky’s analysis of demotion-only update rules reviewed in section 4.2. In section 5.4, I will present an alternative very different proof, that is based on a simple strategy to readapt to the case of OT various results on on-line algorithms for linear classification. Chapter 8 will explore further theoretical consequences of the latter line of analysis. In particular, it will make the point that linear OT has no computational advantage over standard OT, contrary to what has been suggested in the recent literature. Both proofs generalize to a number of variants of the specific promotion-demotion rule of section 5.1. Some of these variants are presented in sections 5.3 and 5.5.

5.1 A principled promotion-demotion update rule

In this section, I construct an update rule for the on-line algorithm (116) that performs both promotion and demotion. The construction is split up into a few steps, with each step motivated by heuristic considerations on the ranking logic of OT. Sections 5.2 and 5.4 will then study this rule in great detail and from different perspectives. Let \mathbf{a} be the comparative row currently fed in step 1 of the on-line algorithm (116); assume that the current ranking vector θ^{old} is not OT-compatible with this row \mathbf{a} , so that action needs to be taken. To get started, assume that this comparative row contains a unique entry equal to L and a unique entry equal to W, as in (120). This is a particularly simple case, because we know (just by virtue of the OT-compatibility of the input tableau) that the constraint corresponding to the unique W of row \mathbf{a} can and must be ranked above the constraint corresponding to the unique L, irrespectively of the rest of the input comparative tableau. In other words, the information contained in this comparative row alone allows us to confidently promote the constraint corresponding to the unique W. Thus, in this case we can promote the unique winner-preferring constraint by the same amount (say 1) we demote the unique loser-preferring constraint, as in the case of Boersma’s update rule (78).

$$(120) \quad \mathbf{a} \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & \mathbf{W} & \dots & \mathbf{L} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Consider next the case where the current comparative row \mathbf{a} still contains a unique entry equal to L but now contains multiple entries equal to W. For concreteness, consider for example the case of the comparative row \mathbf{a} in (121), which contains the two W's corresponding to the two constraints C_h and C_k . This case is much more delicate than the preceding case. As noted in the passage quoted in (81) from Tesar and Smolensky (1998), the problem here is that the comparative row \mathbf{a} by itself does not provide information on which one of the two winner-preferring constraints C_h and C_k we should promote. We crucially need information concerning the entire tableau. For instance, because of the row \mathbf{a}' in the comparative tableau (121), only the winner-preferring constraint C_h will eventually be ranked above the loser-preferring constraint C_ℓ , not the constraint C_k . But given the row \mathbf{a} , with no knowledge of other rows such as \mathbf{a}' , there seems to be no way to choose in a principled way which one between C_h and C_k we should promote.

$$(121) \quad \begin{array}{c} \mathbf{a} \\ \mathbf{a}' \end{array} \left[\begin{array}{ccc} & C_h & C_k & & C_\ell & & \\ & \vdots & \vdots & & \vdots & & \\ \dots & W & W & \dots & L & \dots & \\ & & L & & W & & \\ & \vdots & \vdots & & \vdots & & \end{array} \right]$$

Boersma's promotion-demotion update rule (78) treats the case in (121) analogously to the preceding case in (120): all winner-preferring constraints get promoted by the same amount (say 1), no matter whether they appear in a simple row with a unique winner-preferring constraint as in (120) or in a challenging row with multiple winner-preferring constraints as in (121). This does not look like a good idea though: we want our update rule to be sensitive to the intrinsic logic of OT, namely to be sensitive to the crucial difference just discussed between the two cases (120) and (121). The intuition I would like to put forward is as follows: in the simple case of the comparative row \mathbf{a} in (120) with a unique winner-preferring constraint, we can confidently promote that unique winner-preferring constraint by 1; in the challenging case of the comparative row \mathbf{a} in (121) with two winner-preferring constraints, we should be cautious and split our confidence between the two winner-preferring constraints, by promoting each one just by 1/2. In the general case, if the current comparative row \mathbf{a} contains many entries equal to W, say a total of $w(\mathbf{a})$ entries equal to W, then the uncertainty depends on the size of $w(\mathbf{a})$, and we should thus promote each winner-preferring constraint just by $1/w(\mathbf{a})$. In conclusion, I suggest the new rule that updates the current ranking vector θ^{old} to the new ranking vector θ^{new} in response to a comparative row \mathbf{a} as in (122a), described in words in (122b). This update rule (122) performs both promotion of all winner-preferring constraints and demotion of the unique loser-preferring constraint; thus, it is a *promotion-demotion* update rule. This update rule (122) demotes by 1 but promotes by $1/w(\mathbf{a})$, where $w(\mathbf{a})$ is the number of entries equal to W in the current comparative row \mathbf{a} ; thus, this update rule is *asymmetric* (at least in the case of comparative rows that have multiple entries equal to W).

(122) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} and the latter contains a unique entry equal to L:

$$\text{a. } \theta_k^{\text{new}} \doteq \begin{cases} \theta_k^{\text{old}} + \frac{1}{w(\mathbf{a})} & \text{if } k \in W(\mathbf{a}) \\ \theta_k^{\text{old}} - 1 & \text{if } k = L(\mathbf{a}) \\ \theta_k^{\text{old}} & \text{otherwise} \end{cases}$$

- b. i. Demote the unique currently undominated loser-preferring constraint by 1;
 ii. promote all winner-preferring constraints by $1/w(\mathbf{a})$, where $w(\mathbf{a})$ is the number of winner preferring constraints w.r.t. the comparative row \mathbf{a} .

Of course, we have the same update rule if we multiply both the promotion amount $1/w(\mathbf{a})$ and the demotion amount 1 by the same positive constant. In particular, by multiplying both the promotion and the demotion amounts by the total number $w(\mathbf{a})$ of winner-preferring constraints, we get the equivalent update rule (123a), described in words in (123b). Thus restated, we see another intuitive justification for this update rule. Upon update triggered by a comparative row \mathbf{a} according to the

update rule (123), the highest ranked winner-preferring constraint raises by 1 and the loser-preferring constraint sinks by the total number $w(\mathbf{a})$ of winner preferring constraints. Thus, upon update triggered by \mathbf{a} , the separation between the loser-preferring constraint and the highest ranked winner-preferring constraint increases by $w(\mathbf{a}) + 1$. If the comparative row contains a small number $w(\mathbf{a})$ of entries equal to W , then the separation between the loser-preferring constraint and the highest ranked winner-preferring constraint increases only by the small amount $w(\mathbf{a}) + 1$. Since this is a small separation, it will be eaten up fast by other competing rows. Thus this row will trigger many updates over time. This is good, because we can trust the promotion triggered by a row with few entries equal to W . If instead the comparative row \mathbf{a} contains a large number $w(\mathbf{a})$ of entries equal to W , then the separation between the loser-preferring constraint and the highest winner-preferring constraint will increase by the larger amount $w(\mathbf{a}) + 1$ after a single update by that row. Since this is a large separation, it will take a while for other competing rows to eat up that large separation. Thus, this row will trigger few updates. This is good, because some of the promotions triggered by this row should not have happened. Henceforth, I will use the variant in (123), since it has the additional property that the components of the current ranking vector are integers at every time (provided of course that the initial ranking vector has integer components).

(123) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} and the latter contains a unique entry equal to L :

$$\text{a. } \theta_k^{\text{new}} \doteq \begin{cases} \theta_k^{\text{old}} + 1 & \text{if } k \in W(\mathbf{a}) \\ \theta_k^{\text{old}} - w(\mathbf{a}) & \text{if } k = L(\mathbf{a}) \\ \theta_k^{\text{old}} & \text{otherwise} \end{cases}$$

- b. i. Promote by 1 all winner-preferring constraints;
 ii. demote the unique currently undominated loser-preferring constraint by $w(\mathbf{a})$, where $w(\mathbf{a})$ is the number of winner preferring constraints w.r.t. the comparative row \mathbf{a} .

So far, I have only considered the case where the current comparative row \mathbf{a} has a unique L . If that comparative row is not OT-compatible with the current ranking vector, then its unique loser-preferring constraint must be currently undominated, namely it cannot be ranked below the currently top ranked winner-preferring constraint. We now need to consider the case where the current comparative row has multiple entries equal to L . In this case, some of them might be currently undominated and some others might not be. Consistent with the discussion of the maximal update rule (71) in subsection 4.2.5, we wish to only demote the currently undominated loser-preferring constraints. To this end, let $L(\mathbf{a}, \theta)$ be the set of currently undominated loser-preferring constraints, namely the set of loser-preferring constraints that are ranked by $\theta = (\theta_1, \dots, \theta_n)$ above the top-ranked winner-preferring constraint, as defined in (50) and repeated in (124); furthermore, let me denote by $\ell(\mathbf{a}, \theta)$ the total number of undominated loser-preferring constraints.

$$(124) \quad L(\mathbf{a}, \theta) \doteq \left\{ k \in \{1, \dots, n\} \mid a_k = L, \theta_k \geq \max_{h \in W(\mathbf{a})} \theta_h \right\}$$

$$\ell(\mathbf{a}, \theta) \doteq \text{the cardinality of the set } L(\mathbf{a}, \theta)$$

If only one loser-preferring constraint is currently undominated (namely $L(\mathbf{a}, \theta)$ is a singleton), then we can of course use again the very same update rule (123), where in this case we of course replace $L(\mathbf{a})$ by $L(\mathbf{a}, \theta)$. What if there are more than one currently undominated loser-preferring constraints, namely $L(\mathbf{a}, \theta)$ is not a singleton? For concreteness, suppose that there are two currently undominated loser-preferring constraints C' and C'' , as in the row \mathbf{a} in (125). As noted in the discussion preceding claim 2, we get an OT-equivalent comparative tableau if we split up that comparative row \mathbf{a} into two rows \mathbf{a}' and \mathbf{a}'' such that each of them retains only one of those two L 's of the original row \mathbf{a} , while the other one gets replaced by an E .

$$(125) \quad \mathbf{a} \begin{array}{cccccc} & & & C' & C'' & \\ & & & \vdots & \vdots & \\ & & & \vdots & \vdots & \\ \dots & W & \dots & L & L & \dots \\ & \vdots & & \vdots & \vdots & \\ & \vdots & & \vdots & \vdots & \end{array} \Rightarrow \begin{array}{cccccc} & & & C' & C'' & \\ & & & \vdots & \vdots & \\ & & & \vdots & \vdots & \\ a' & \dots & W & \dots & L & \dots \\ a'' & \dots & W & \dots & & L & \dots \\ & \vdots & & \vdots & \vdots & \\ & \vdots & & \vdots & \vdots & \end{array}$$

Because of the OT-equivalence between the two comparative tableaux in (125), it makes sense to construe the update triggered by the row \mathbf{a} as the sequence of the two updates triggered by the two rows \mathbf{a}' and \mathbf{a}'' . Furthermore, since the latter two rows contain a single L, then we can update in response to each of these two rows using the update rule (123) devised above for the case of comparative rows that contain a unique entry equal to L. Note that the two derived rows \mathbf{a}' and \mathbf{a}'' both have the same number $w(\mathbf{a})$ of entries equal to W as the original row \mathbf{a} . Thus, upon update by \mathbf{a}' according to (123), the loser-preferring constraint C' gets demoted by $w(\mathbf{a})$ and all winner-preferring constraints get promoted by 1. And upon subsequent update triggered by \mathbf{a}'' , the loser-preferring constraint C'' gets demoted by $w(\mathbf{a})$ too and all winner-preferring constraints get promoted once more by 1. In the end, each one of the undominated loser-preferring constraints of \mathbf{a} gets demoted by the total number $w(\mathbf{a})$ of winner-preferring constraints. Furthermore, each one of the winner-preferring constraints gets promoted by 1 for as many times as there are undominated loser-preferring constraints. Equivalently, each one of the winner-preferring constraints gets promoted by the total number $\ell(\mathbf{a}, \theta^{\text{old}})$ of currently undominated loser-preferring constraints. These heuristic considerations suggest the new rule that updates the current ranking vector θ^{old} to the new ranking vector θ^{new} in response to a comparative row \mathbf{a} as in (126a), described in words in (126b).

(126) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} :

$$\text{a. } \theta_k^{\text{new}} \doteq \begin{cases} \theta_k^{\text{old}} + \ell(\mathbf{a}, \theta^{\text{old}}) & \text{if } k \in W(\mathbf{a}) \\ \theta_k^{\text{old}} - w(\mathbf{a}) & \text{if } k \in L(\mathbf{a}, \theta^{\text{old}}) \\ \theta_k^{\text{old}} & \text{otherwise} \end{cases}$$

- b. i. Promote all winner-preferring constraints by the number $\ell(\mathbf{a}, \theta^{\text{old}})$ of currently undominated loser-preferring constraints;
 ii. demote all currently undominated loser-preferring constraints by the number $w(\mathbf{a})$ of winner-preferring constraints.

Of course, the update rule (126) fits into the additive scheme $\theta^{\text{new}} = \theta^{\text{old}} + \bar{\mathbf{a}}$ in (46) once the update vector $\bar{\mathbf{a}} = (\bar{a}_1, \dots, \bar{a}_n)$, corresponding to a current comparative row \mathbf{a} and a current ranking vector θ^{old} , is defined as in (127): if C_k is a winner-preferring constraint, then the corresponding entry \bar{a}_k of the update vector is the number $\ell(\mathbf{a}, \theta)$ of undominated loser-preferring constraints, as defined in (124); if C_k is an undominated loser-preferring constraint, then the corresponding entry \bar{a}_k of the update vector is the opposite of the number $w(\mathbf{a})$ of winner-preferring constraints; all other entries of the update vector are null. Note that the update vector $\bar{\mathbf{a}}$ depends not only on the current comparative row \mathbf{a} but also on the current ranking vector θ^{old} (even though the notation $\bar{\mathbf{a}}$ does not transparently reflect the dependence on θ^{old}).

$$(127) \quad \bar{a}_k = \begin{cases} \ell(\mathbf{a}, \theta^{\text{old}}) & \text{if } k \in W(\mathbf{a}) \\ -w(\mathbf{a}) & \text{if } k \in L(\mathbf{a}, \theta^{\text{old}}) \\ 0 & \text{otherwise} \end{cases}$$

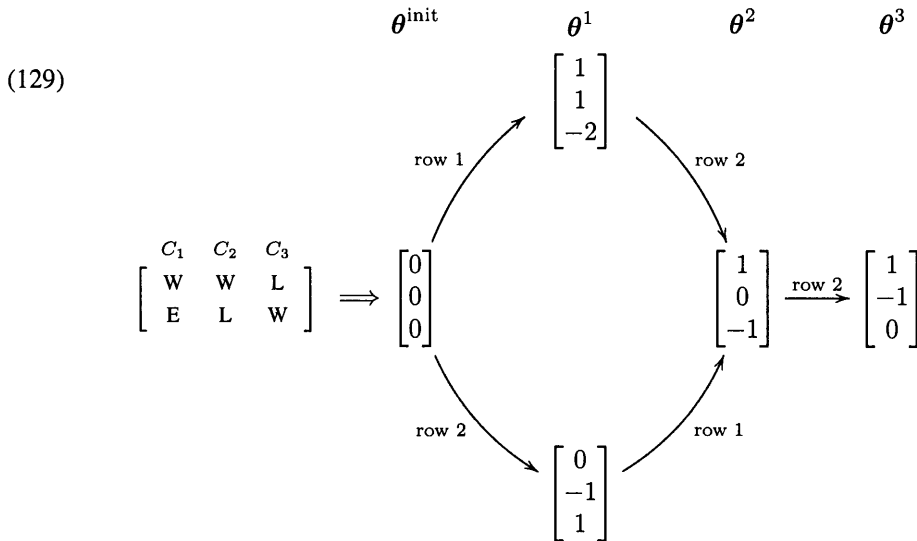
Of course, I can again divide both the promotion amount $\ell(\mathbf{a}, \theta^{\text{old}})$ and the demotion amount $w(\mathbf{a})$ by the same positive quantity $w(\mathbf{a}) \cdot \ell(\mathbf{a}, \theta)$ and get the equivalent update rule in (128), where the promotion amount only depends on the number $w(\mathbf{a})$ of winner-preferring constraints and the demotion amount only depends on the number $\ell(\mathbf{a}, \theta^{\text{old}})$ of currently undominated loser-preferring constraints. Again, I will usually prefer the variant (126) over (128), since the former has the additional property that the components of the current ranking vector are always integral (provided that the initial ranking vector has integer components).

(128) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} :

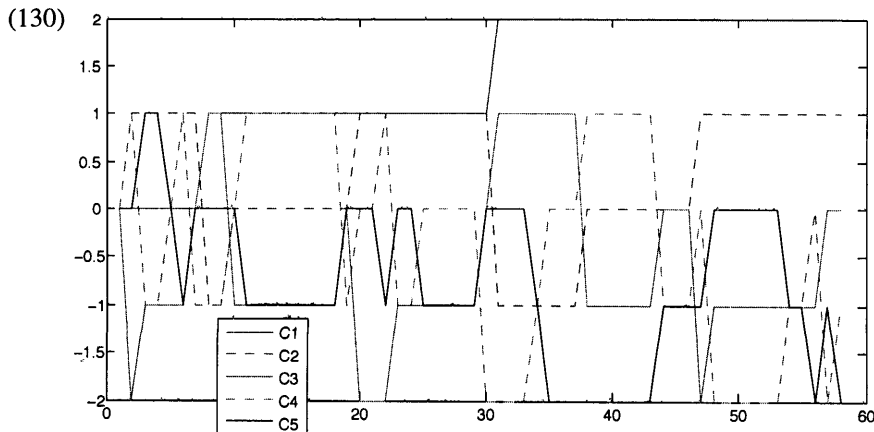
$$a. \theta_k^{\text{new}} \doteq \begin{cases} \theta_k^{\text{old}} + \frac{1}{w(\mathbf{a})} & \text{if } k \in W(\mathbf{a}) \\ \theta_k^{\text{old}} - \frac{1}{\ell(\mathbf{a}, \theta^{\text{old}})} & \text{if } k \in L(\mathbf{a}, \theta^{\text{old}}) \\ \theta_k^{\text{old}} & \text{otherwise} \end{cases}$$

- b. i. Promote all winner-prefering constraints by the inverse of the number $w(\mathbf{a})$ of winner-prefering constraints;
- ii. demote all currently undominated loser-prefering constraints by the inverse of the number $\ell(\mathbf{a}, \theta^{\text{old}})$ of currently undominated loser-prefering constraints.

To illustrate, the behavior of the on-line algorithm (116) with the new promotion-demotion update rule (126) run on the input comparative tableau (15) is described in the diagram (129). Note the non-monotonic dynamics of the ranking values of the two constraints C_1 and C_2 over time, which is the hallmark of update rules that perform promotion too.



The graph in (130) shows the dynamics over time of the components of the ranking vector entertained by the on-line algorithm (116) with the promotion-demotion update rule (126) run on Pater’s comparative tableau (82) starting from the null initial vector. The algorithm does converge to the right ranking vector, contrary to the case of Boersma’s (1997) promotion-demotion update rule (78). Yet, the resulting dynamics is rather complicated, with a large degree of oscillations.



The preceding heuristic considerations motivate the new promotion-demotion update rule (126). Sections 5.2 and 5.4 offer two different proofs that the on-line algorithm (116) with the promotion-demotion update rule (126) satisfies condition (117), namely can only perform a finite number of updates. These two different proofs will lead to various other families of promotion-demotion update rules, presented in sections 5.3 and 5.5. We thus have a full arsenal of update rules that yield a number of variants of the OT on-line model of the acquisition of phonology presented in section 4.2.

5.2 First proof of finite time convergence

In this section, I present a first proof that condition (117) holds for the promotion-demotion update rule (126), namely that the corresponding on-line algorithm (116) can perform only a finite number of updates. Throughout this section, I concentrate on the case where the initial ranking vector is the null vector; the extension to an arbitrary ranking vector is straightforward. The proof presented in this section is close in spirit to Tesar and Smolensky's analysis of the minimal demotion-only update rule (51), as reviewed in subsections 4.2.2-4.2.4. This proof has four steps.

5.2.1 First step

The first step consists of claim 12. This claim is a straightforward generalization of Tesar and Smolensky's claim 5 from section 4.2.1

Claim 12 Consider a comparative tableau A and a ranking \gg OT-compatible with A . Without loss of generality, assume that this ranking is $C_1 \gg C_2 \gg \dots \gg C_n$ (otherwise, just relabel the constraints). For every $k = 1, \dots, n$, let $dec(k)$ be the set of rows whose \gg -highest winner preferring constraint is C_k , as defined in (16b) and repeated in (131).

$$(131) \quad dec(k) \doteq \left\{ \mathbf{a} = (a_1, \dots, a_n) \mid \begin{array}{l} a_k = W, \\ a_1 = \dots = a_{k-1} = E \end{array} \right\}$$

The ranking vector $\theta^t = (\theta_1^t, \dots, \theta_n^t)$ entertained at any time t by the on-line algorithm (116) run on the input comparative tableau A with the update rule (126) starting from the null initial vector satisfies condition (132).

$$(132) \quad \begin{array}{l} \theta_1^t \geq 0 \\ \theta_k^t \geq - \sum_{h=1}^{k-1} \max_{\mathbf{a} \in dec(h)} w(\mathbf{a}), \quad k = 2, \dots, n \end{array}$$

Proof. Condition (132) trivially holds for $k = 1$ at any time t , since the top ranked constraint C_1 can never be a loser-preferrer for any row, and thus its ranking value θ_1 can only increase w.r.t. its initial value $\theta_1^{\text{init}} = 0$. Let me prove that property (132) holds for $k = 2, \dots, n$ by induction on t . Claim (132) trivially holds for every $k = 2, \dots, n$ at time $t = 0$, because of the assumption that $\theta^{\text{init}} = \mathbf{0}$. Assume that claim (132) holds at time $t - 1$ for every $k = 2, \dots, n$ and let me prove that it then holds at time t for any given $k = 2, \dots, n$. Suppose that at time $t - 1$ the algorithm gets the row \mathbf{a} of the input comparative tableau. There are three possible cases to be considered, listed in (133). If either case I or case II holds, then claim (132) trivially holds for k at time t , because of the inductive hypothesis that it held for k at time $t - 1$ together with the fact that θ_k^t is either equal to (in case I) or larger than (in case II) θ_k^{t-1} .

$$(133) \quad \theta_k^t = \begin{cases} \theta_k^{t-1} & \text{if } \theta^{t-1} \text{ is OT-compatible with } \mathbf{a} \text{ or } a_k = E & \text{(case I)} \\ \theta_k^{t-1} + 1 & \text{if } \theta^{t-1} \text{ is not OT-compatible with } \mathbf{a} \text{ and } a_k = W & \text{(case II)} \\ \theta_k^{t-1} - w(\mathbf{a}) & \text{if } \theta^{t-1} \text{ is not OT-compatible with } \mathbf{a} \text{ and } a_k = L & \text{(case III)} \end{cases}$$

Thus, I only have to consider case III. Let C_h be the \gg -highest ranked constraint that has a W in the given row \mathbf{a} , namely $h \in \{1, \dots, n\}$ is such that $\mathbf{a} \in dec(h)$. Since the k th entry of the row \mathbf{a} is an L , then it must be $h < k$, otherwise the row \mathbf{a} would not be OT-compatible with the ranking

$C_1 \gg \dots \gg C_h \gg \dots \gg C_k \gg \dots \gg C_n$. Since the row \mathbf{a} is not OT-compatible with the current ranking vector θ^{t-1} , then C_h must be currently ranked no higher than C_k , namely $\theta_h^{t-1} \leq \theta_k^{t-1}$. The component θ_k^t can thus be lower bound as in (134).

$$\begin{aligned}
 (134) \quad \theta_k^t &\stackrel{(a)}{=} \theta_k^{t-1} - w(\mathbf{a}) \\
 &\stackrel{(b)}{\geq} \theta_h^{t-1} - w(\mathbf{a}) \\
 &\stackrel{(c)}{\geq} \theta_h^{t-1} - \max_{\mathbf{a} \in \text{dec}(h)} w(\mathbf{a}) \\
 &\stackrel{(d)}{\geq} - \sum_{\ell=1}^{h-1} \max_{\mathbf{a} \in \text{dec}(\ell)} w(\mathbf{a}) - \max_{\mathbf{a} \in \text{dec}(h)} w(\mathbf{a}) \\
 &= - \sum_{\ell=1}^h \max_{\mathbf{a} \in \text{dec}(\ell)} w(\mathbf{a}) \\
 &\stackrel{(e)}{\geq} - \sum_{\ell=1}^{k-1} \max_{\mathbf{a} \in \text{dec}(\ell)} w(\mathbf{a})
 \end{aligned}$$

Here, I have reasoned as follows: in step (a), I have used the hypothesis that we are in case III; in step (b), I have used the fact that $\theta_h^{t-1} \leq \theta_k^{t-1}$, as noted above; in step (c), I have used the fact that $\mathbf{a} \in \text{dec}(h)$, as noted above; in step (d), I have used the inductive hypothesis that (132) holds at time $t-1$; in step (e), I have used the fact that $h < k$. ■

5.2.2 Second step

The preceding claim 12 provides a lower bound on the components of the ranking vector entertained by the algorithm at a generic time, thus ensuring that these components cannot become arbitrarily small. The next step of the reasoning is the following claim 13, that provides an upper bound on the components of the ranking vector entertained by the algorithm at a generic time, thus ensuring that these components cannot become too large neither. Recall that no upper bound holds for the case of Boersma's promotion-demotion update rule (78): as shown by Pater's counterexample (82), the components of the ranking vector can increase indefinitely with time when Boersma's update rule is used. The following claim is thus important, because it ensures that this cannot happen in the case of the new promotion-demotion update rule (126). The idea used in the proof of the following claim is rather simple: since by claim 12 the components of the ranking vector cannot become too small, then they cannot become too large either, since at any given time they must add up to zero.

Claim 13 The ranking vector $\theta = (\theta_1, \dots, \theta_n)$ entertained at an arbitrary time by the on-line algorithm (116) run on an input OT-compatible comparative tableau \mathbf{A} with the update rule (126) starting from the null initial vector satisfies condition (135) for every $k = 1, \dots, n$.

$$(135) \quad \theta_k \leq \frac{n^3}{2}$$

Proof. Let me start by showing that the components of the ranking vector $\theta = (\theta_1, \dots, \theta_n)$ entertained by the algorithm at a generic time add up to zero, as stated in (136).

$$(136) \quad \sum_{h=1}^n \theta_h = 0$$

The claim trivially holds for the initial null ranking vector θ^{init} . Thus, it is sufficient to show that, if the components of the current ranking vector $\theta^{\text{old}} = (\theta_1^{\text{old}}, \dots, \theta_n^{\text{old}})$ add up to zero, then the components of the ranking vector $\theta^{\text{new}} = (\theta_1^{\text{new}}, \dots, \theta_n^{\text{new}})$ obtained by update (126) add up to zero too. Since the update rule (126) fits into the additive scheme $\theta^{\text{new}} = \theta^{\text{old}} + \bar{\mathbf{a}}$ in (46), then it is in turn sufficient to show that the components of the update vector $\bar{\mathbf{a}} = (\bar{a}_1, \dots, \bar{a}_n)$ in (127) that defined the update rule (126) sum up to zero. This is shown by the chain of identities in (137). Here,

I have reasoned as follows: in step (a), I have split up the set $\{1, \dots, h\}$ that h runs over into the three sets $W(\mathbf{a})$, $L(\mathbf{a}, \theta^{\text{old}})$ and their complement; in step (b), I have used the definition (127) of the update rule.

$$\begin{aligned}
 (137) \quad \sum_{h=1}^n \bar{a}_h &\stackrel{(a)}{=} \sum_{h \in W(\mathbf{a})} \bar{a}_h + \sum_{h \in L(\mathbf{a}, \theta^{\text{old}})} \bar{a}_h + \sum_{h \notin W(\mathbf{a}) \cup L(\mathbf{a}, \theta^{\text{old}})} \bar{a}_h \\
 &\stackrel{(b)}{=} \sum_{h \in W(\mathbf{a})} \ell(\mathbf{a}, \theta^{\text{old}}) - \sum_{h \in L(\mathbf{a}, \theta^{\text{old}})} w(\mathbf{a}) \\
 &= w(\mathbf{a})\ell(\mathbf{a}, \theta^{\text{old}}) - w(\mathbf{a})\ell(\mathbf{a}, \theta^{\text{old}}) \\
 &= 0
 \end{aligned}$$

The identity (136) says in particular that the components of the current ranking vector θ cannot all have the same sign, namely that some of them must be positive and some of them must be negative. Let P and N be the set of indices corresponding to positive and negative components of θ respectively, as defined in (138). Note that some (possibly all) components of θ could of course be null, and thus belong neither to P nor to N .

$$\begin{aligned}
 (138) \quad P &= \{k \in \{1, \dots, n\} \mid \theta_k > 0\} \\
 N &= \{k \in \{1, \dots, n\} \mid \theta_k < 0\}
 \end{aligned}$$

If $k \in N$, then $\theta_k < 0$, and thus the bound in (135) holds a fortiori. Thus, I only need to prove the bound in (135) for an arbitrary $\bar{k} \in P$. The proof consists of the chain of inequalities in (139).

$$\begin{aligned}
 (139) \quad \theta_{\bar{k}} &\stackrel{(a)}{\leq} \sum_{k \in P} \theta_k \\
 &\stackrel{(b)}{=} - \sum_{k \in N} \theta_k \\
 &\stackrel{(c)}{\leq} \sum_{k \in N} \sum_{h=1}^{k-1} \max_{\mathbf{a} \in \text{dec}(h)} w(\mathbf{a}) \\
 &\stackrel{(d)}{\leq} \sum_{k \in N} \sum_{h=1}^{k-1} n \\
 &= n \sum_{k \in N} (k-1) \\
 &\stackrel{(e)}{\leq} n \sum_{k=2}^n (k-1) \\
 &= n \left(\frac{n^2}{2} - \frac{3}{2}n + 1 \right)
 \end{aligned}$$

Here, I have reasoned as follows: in step (a), I have used the fact that $\bar{k} \in P$ and that $\theta_k \geq 0$ for every $k \in P$; in step (b), I have used the identity (136), which indeed entails that $\sum_{h \in P} \theta_h = - \sum_{h \in N} \theta_h$; in step (c), I have used the inequality (132), under the assumption that the input comparative tableau is OT-compatible with the ranking $C_1 \gg C_2 \gg \dots \gg C_n$ (otherwise, just relabel the constraints); in step (d), I have noted that $w(\mathbf{a}) \leq n$; in step (e), I have used the fact that N can contain at most $n - 1$ components, since at least one component must be positive, in order for the identity (136) to hold. ■

5.2.3 Third step

The third step of the proof consists of the following claim 14. This claim says that, once the current ranking vector θ is updated to a new ranking vector θ' , there is no way for the algorithm to loop back to that ranking vector θ . In other words, once a ranking vector is deemed unsuitable, it is never considered again.

Claim 14 The sequence of ranking vectors entertained by the on-line algorithm (116) with the update rule (126) run on an input OT-compatible comparative tableau starting from the null initial vector cannot contain a subsequence such as (140), whereby the same ranking vector θ is entertained twice but with some other ranking vector $\theta' \neq \theta$ entertained in between.

$$(140) \quad \dots \rightarrow \theta \rightarrow \dots \rightarrow \theta' \rightarrow \dots \rightarrow \theta \rightarrow \dots$$

Proof. Let $\theta^t = (\theta_1^t, \dots, \theta_n^t) \in \mathbb{R}^n$ be the ranking vector entertained by the algorithm immediately after the t th update; thus θ^1 is the ranking vector entertained by the algorithm after the first update, θ^2 is the ranking vector entertained by the algorithm after the second update, and so on. I have to prove the implication (141).

$$(141) \quad \theta^t = \theta^{t'} \implies t = t'$$

Let $\alpha^t = (\alpha_1^t, \dots, \alpha_m^t) \in \mathbb{N}^m$ be the dual vector after the t th update, namely α_i^t is the number of the first t updates that were triggered by the i th row of the input comparative tableau. As noted in (91), the total number of updates t is equal to the number of updates α_1^t triggered by the first row plus the number of updates α_2^t triggered by the second row and so on, namely $t = \sum_{i=1}^m \alpha_i^t$. Analogously, let $\alpha^{t'} = (\alpha_1^{t'}, \dots, \alpha_m^{t'}) \in \mathbb{N}^m$ be the dual vector after the t' th update, so that $t' = \sum_{i=1}^m \alpha_i^{t'}$. Thus, the implication (141) is equivalent to the implication (142).

$$(142) \quad \theta^t = \theta^{t'} \implies \sum_{i=1}^m \alpha_i^t = \sum_{i=1}^m \alpha_i^{t'}$$

Assume without loss of generality that $t \leq t'$. Note then that $\sum_{i=1}^m \alpha_i^t = \sum_{i=1}^m \alpha_i^{t'}$ iff $\alpha^t = \alpha^{t'}$, since $\alpha^t \leq \alpha^{t'}$ because of the fact that $t \leq t'$ and that the dual variables are nondecreasing over time. Thus, the implication (142) is equivalent to the implication (143), which says that primal identity entails dual identity.¹ Let me thus show that this implication (142) does indeed hold.

$$(143) \quad \theta^t = \theta^{t'} \implies \alpha^t = \alpha^{t'}$$

Claim 1 ensures that there exists an integer $d \leq n$ such that, by relabeling the constraints and properly reordering the rows and the columns of the input comparative tableau \mathbf{A} , it takes the form in (18), repeated in (144).

$$(144) \quad \begin{array}{c} \begin{array}{cccccc} C_1 & C_2 & \dots & C_d & C_{d+1} & \dots & C_n \end{array} \\ \begin{array}{l} dec(C_1) \\ dec(C_2) \\ \vdots \\ dec(C_d) \end{array} \end{array} \left[\begin{array}{cccccc} W & & & & & & \\ | & \dots & \dots & \dots & \dots & \dots & \dots \\ W & & & & & & \\ E & W & & & & & \\ | & | & \dots & \dots & \dots & \dots & \dots \\ E & W & & & & & \\ E & E & & & & & \\ \vdots & & & & & & \\ | & | & \dots & \dots & \dots & \dots & \dots \\ E & E & & & & & \\ E & E & & W & & & \\ | & E & \dots & | & \dots & \dots & \dots \\ E & E & & W & & & \end{array} \right]$$

The implication in (143) can thus be made more explicit as in (145), by partitioning the components of the dual vectors into the d sets $dec(C_1), \dots, dec(C_d)$. I will now prove the implication (145) by induction on k .

$$(145) \quad \theta^t = \theta^{t'} \implies \alpha_i^t = \alpha_i^{t'} \text{ for every } i \in dec(C_k) \text{ for every } k = 1, \dots, d$$

¹Note that this does not mean that in the general case we can reconstruct the dual vector from the primal vector. The validity of the implication (143) crucially rests on the hypothesis that we are considering two ranking vectors that belong to the same run, so that one of the two is obtained starting from the other.

To simplify the exposition, let me first present the proof under the simplificatory hypothesis that every row of the input comparative tableau (144) contains one and exactly one entry equal to L; then, I will show how the reasoning used in this special case can be extended to the general case. Under this simplificatory hypothesis, the general update rule (126) reduces to the special case (123). And the corresponding update vector \bar{a} in (127) only depends on the current comparative row, not on the current ranking vector, as made explicit in (146). Thus, given an input comparative tableau A , I can consider the numerical matrix \bar{A} obtained by organizing one underneath the other the update vectors corresponding to the comparative rows of A , as was illustrated in 4.2.6.2. Let $\bar{a}_{i,k}$ be the entry of the matrix \bar{A} thus obtained that sits in the i th row and the k th column.²

$$(146) \quad \mathbf{a} = [a_1 \quad \dots \quad a_k \quad \dots \quad a_n] \quad \Longrightarrow \quad \bar{\mathbf{a}} = [\bar{a}_1 \quad \dots \quad \bar{a}_k \quad \dots \quad \bar{a}_n]$$

$$a_k \in \{L, E, W\} \quad \bar{a}_k \doteq \begin{cases} +1 & \text{if } a_k = W \\ 0 & \text{if } a_k = E \\ -w(\mathbf{a}) & \text{if } a_k = L \end{cases}$$

To establish the base case of the induction, let me show that the implication (145) holds for $k = 1$, namely that $\theta^t = \theta^{t'}$ entails that $\alpha_i^t = \alpha_i^{t'}$ for every $i \in \text{dec}(C_1)$. The proof consists of the chain of implications in (147). Here, I have reasoned as follows. In step (a), I have expressed the primal variables θ_1^t and $\theta_1^{t'}$ in terms of the corresponding dual variables α^t and $\alpha^{t'}$, reasoning as for the identity (95) in subsection 4.2.6. In other words, I have reasoned as follows. The ranking value of the constraint C_1 starts out at 0. The update rule (123) modifies the ranking value of constraint C_1 only when update is triggered by a row in $\text{dec}(C_1)$. Furthermore, each time a row in $\text{dec}(C_1)$ triggers an update, the ranking value of constraint C_1 gets incremented by 1. Thus, the ranking value θ_1^t of constraint C_1 after the t th update is equal to the number of those t updates that were triggered by a row in $\text{dec}(C_1)$. Equivalently, θ_1^t is equal to the sum of α_i^t over all $i \in \text{dec}(C_1)$. Analogous considerations hold of course for $\theta_1^{t'}$. In step (b), I have used the fact that $\alpha^t \leq \alpha^{t'}$, since $t \leq t'$ by hypothesis and furthermore the dual variables are nondecreasing over time.

$$(147) \quad \theta^t = \theta^{t'} \quad \Longrightarrow \quad \theta_1^t = \theta_1^{t'}$$

$$\stackrel{(a)}{\Longrightarrow} \quad \underbrace{\sum_{i \in \text{dec}(C_1)} \alpha_i^t}_{\theta_1^t} = \underbrace{\sum_{i \in \text{dec}(C_1)} \alpha_i^{t'}}_{\theta_1^{t'}}$$

$$\stackrel{(b)}{\Longrightarrow} \quad \alpha_i^t = \alpha_i^{t'} \text{ for every } i \in \text{dec}(C_1)$$

Before I turn to the general case, let me show that the implication (145) holds for $k = 2$, namely let me show that $\theta^t = \theta^{t'}$ entails that $\alpha_i^t = \alpha_i^{t'}$ for every $i \in \text{dec}(C_2)$. The proof consists of the chain of implications in (148), which is completely analogous to the one in (147). Here, I have reasoned as follows. In step (a), I have expressed the primal variable θ_2^t after the t th update in terms of the corresponding dual variables α^t and the coefficients $\bar{a}_{i,2}$ through (95). In other words, I have reasoned as follows. The ranking value of the constraint C_2 starts out at 0. The update rule (123)

²The high-level idea of the proof of the implication (143) can now be made explicit as follows, using standard facts from Linear Algebra. As noted in footnote 18, the relation (i) holds: the primal vector θ^t at time t is the row-by-column product between the transpose of the matrix \bar{A} , obtained by organizing the update vectors one underneath the other, and the corresponding dual vector α^t .

$$(i) \quad \text{a. } \theta^t = \bar{A}^T \alpha^t$$

$$\text{b. } \alpha^t = (\bar{A}^{-1})^T \theta^t$$

Suppose now that the input comparative tableau in (144) had the property that $d = n$ and furthermore that $\text{dec}(C_k)$ was a singleton for every constraint C_k . Then, the corresponding matrix \bar{A} of update vectors would be a square triangular matrix. A triangular matrix is invertible. Thus, we could invert (i) into (ii). The relation in (ii) says that the primal vector determines the dual vector, and thus in particular it entails (145). In the general case, the numerical matrix \bar{A} is not even a square matrix, hence it is not invertible and therefore it is not true that the primal vector determines the dual vector. Yet, in the general case, OT-compatibility of a comparative tableau A entails that the corresponding matrix \bar{A} of update vectors is in some sense "close" to a triangular matrix. This ensures that the primal vector determines the dual vector at least if we restrict ourselves to primal and dual vectors in a single run, as stated in (143).

modifies the ranking value of constraint C_2 only when update is triggered by a row in $dec(C_1)$ or in $dec(C_2)$. Furthermore, each time a row in $dec(C_2)$ triggers an update, the ranking value of constraint C_1 gets incremented by 1. And each time a row a_i in $dec(C_1)$ triggers an update, the ranking value of constraint C_1 gets modified by the amount $\bar{a}_{i,2}$. Thus, the ranking value θ_1^t of constraint C_1 after the t th update is equal to the sum of α_i^t over all $i \in dec(C_2)$ plus the sum of α_i^t over all $i \in dec(C_1)$ each multiplied by $\bar{a}_{i,2}$. Analogous considerations hold of course for θ_2^t . In step (b), I have used the inductive hypothesis that $\alpha_i^t = \alpha_i^{t'}$ for every $i \in dec(C_1)$; in step (c), I have used the fact that $\alpha^t \leq \alpha^{t'}$, since $t \leq t'$ by hypothesis and furthermore the dual variables are nondecreasing over time.

$$\begin{aligned}
(148) \quad \theta^t = \theta^{t'} &\implies \theta_2^t = \theta_2^{t'} \\
&\stackrel{(a)}{\implies} \underbrace{\sum_{i \in dec(C_1)} \bar{a}_{i,2} \alpha_i^t + \sum_{i \in dec(C_2)} \alpha_i^t}_{\theta_2^t} = \underbrace{\sum_{i \in dec(C_1)} \bar{a}_{i,2} \alpha_i^{t'} + \sum_{i \in dec(C_2)} \alpha_i^{t'}}_{\theta_2^{t'}} \\
&\stackrel{(b)}{\implies} \sum_{i \in dec(C_2)} \alpha_i^t = \sum_{i \in dec(C_2)} \alpha_i^{t'} \\
&\stackrel{(c)}{\implies} \alpha_i^t = \alpha_i^{t'} \text{ for every } i \in dec(C_k)
\end{aligned}$$

Assume now that $\alpha_i^t = \alpha_i^{t'}$ for every $i \in dec(C_1) \cup \dots \cup dec(C_{k-1})$ for some $k \leq d$ and let me prove the inductive step that $\alpha_i^t = \alpha_i^{t'}$ for every $i \in dec(C_k)$. The proof consists of the chain of implications in (149), which is completely analogous to the ones in (147) and (148). Here, I have reasoned as follows: in step (a), I have expressed the primal variables θ_k^t and $\theta_k^{t'}$ in terms of the corresponding dual variables α^t and $\alpha^{t'}$ and the coefficients $\bar{a}_{i,k}$ through (95); in step (b), I have used the inductive hypothesis that $\alpha_i^t = \alpha_i^{t'}$ for every $i \in dec(C_1) \cup \dots \cup dec(C_{k-1})$; in step (c), I have used the fact that $\alpha^t \leq \alpha^{t'}$, since $t \leq t'$ by hypothesis and furthermore the dual variables are nondecreasing over time.

$$\begin{aligned}
(149) \quad \theta^t = \theta^{t'} &\implies \theta_k^t = \theta_k^{t'} \\
&\stackrel{(a)}{\implies} \underbrace{\sum_{i \in dec(C_1) \cup \dots \cup dec(C_{k-1})} \bar{a}_{i,k} \alpha_i^t + \sum_{i \in dec(C_k)} \alpha_i^t}_{\theta_k^t} \\
&= \underbrace{\sum_{i \in dec(C_1) \cup \dots \cup dec(C_{k-1})} \bar{a}_{i,k} \alpha_i^{t'} + \sum_{i \in dec(C_k)} \alpha_i^{t'}}_{\theta_k^{t'}} \\
&\stackrel{(b)}{\implies} \sum_{i \in dec(C_k)} \alpha_i^t = \sum_{i \in dec(C_k)} \alpha_i^{t'} \\
&\stackrel{(c)}{\implies} \alpha_i^t = \alpha_i^{t'} \text{ for every } i \in dec(C_k)
\end{aligned}$$

To conclude the proof, I need to consider the case where the input comparative tableau \mathbf{A} contains rows with multiple entries equal to L , so that the general update rule (126) does not reduce to the special update rule (123). The crucial difficulty in this case is that the contribution of the i th row of the input tableau to the primal vector depends on the number of currently undominated loser-preferring constraints, and thus cannot be distilled into a unique coefficient $\bar{a}_{i,k}$. To overcome this difficulty, we just need a more careful definition of the dual variables. Suppose that the i th row of the input tableau has ℓ entries equal to L , corresponding to the loser-preferring constraints $C_{i_1}, \dots, C_{i_\ell}$. Let $C_1, \dots, C_j, \dots, C_{2^\ell - 1}$ be all the non empty subsets of the set $\{C_{i_1}, \dots, C_{i_\ell}\}$. For every $j = 1, \dots, 2^\ell - 1$, let $\alpha_{i,j}^t$ be the number of the first t updates that were triggered by the i th row of the input tableau because all and only the loser-preferring constraints in the set C_j were currently undominated. I can thus repeat the preceding reasoning using these refined dual variables $\alpha_{i,j}^t$. ■

5.2.4 Fourth step

I am now ready to conclude my reasoning with a trivial proof of finite time convergence of the on-line algorithm (116) with the promotion-demotion update rule (126).

Claim 15 The on-line algorithm (116) run on an input OT-compatible comparative tableau starting from the null initial vector with the promotion-demotion update rule (126) can only make a finite number of updates in step 3.

Proof. By claims 12 and 13, the search space of the algorithm is a bounded region of \mathbb{R}^n . Furthermore, the algorithm can only consider integral ranking vectors. Thus, the search space of the algorithm is finite. Since the algorithm cannot entertain the same ranking vector twice by claim 14, then the algorithm must converge after a finite number of updates. ■

5.3 Variants

In this section, I discuss under which conditions the reasoning presented in the preceding section extends from the specific promotion-demotion update rule (126) to a general minimal promotion-demotion update rule of the form (46), repeated in (150). Or equivalently, from the update vector (127) to a general update vector $\bar{\mathbf{a}} = (\bar{a}_1, \dots, \bar{a}_n)$.

(150) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} :

$$\text{a. } \begin{bmatrix} \theta_1^{\text{new}} \\ \vdots \\ \theta_k^{\text{new}} \\ \vdots \\ \theta_n^{\text{new}} \end{bmatrix} = \begin{bmatrix} \theta_1^{\text{old}} \\ \vdots \\ \theta_k^{\text{old}} \\ \vdots \\ \theta_n^{\text{old}} \end{bmatrix} + \begin{bmatrix} \bar{a}_1 \\ \vdots \\ \bar{a}_k \\ \vdots \\ \bar{a}_n \end{bmatrix} \quad \text{where } \bar{a}_k \text{ is } \begin{cases} \leq 0 & \text{if } k \in L(\mathbf{a}, \theta^{\text{old}}) \\ \geq 0 & \text{if } k \in W(\mathbf{a}) \\ = 0 & \text{otherwise} \end{cases}$$

|
update vector

- b. If the constraint C_k is winner-preferring, promote it by adding the nonnegative constant $\bar{a}_k \geq 0$; if the constraint C_k is loser-preferring and undominated, demote it by adding the nonpositive constant $\bar{a}_k \leq 0$.

This discussion will allow me to further comment on and clarify the reasoning presented in the preceding section.

5.3.1 First variant

Without loss of generality, assume that the ranking that the input comparative tableau is OT-compatible with is $C_1 \gg C_2 \gg \dots \gg C_n$. For every $h = 1, \dots, n$, for every row $\mathbf{a} \in \text{dec}(C_h)$ of the input comparative tableau and for every update vector $\bar{\mathbf{a}}$ corresponding to \mathbf{a} , let me also write $\bar{\mathbf{a}} \in \text{dec}(C_h)$. Claim 12 straightforwardly extends from the specific promotion-demotion update rule (126) to the general minimal update rule (150) with the bound (132) generalized to (151), without any special restrictions on the update vector $\bar{\mathbf{a}} = (\bar{a}_1, \dots, \bar{a}_n)$.

$$(151) \quad \begin{aligned} \theta_1^i &\geq 0 \\ \theta_k^i &\geq \sum_{h=1}^{k-1} \max_{\bar{\mathbf{a}} \in \text{dec}(C_h)} \min\{\bar{a}_k, 0\}, \quad k = 2, \dots, n \end{aligned}$$

Consider the update vector $\bar{\mathbf{a}} = (\bar{a}_1, \dots, \bar{a}_n)$ defined by $\bar{a}_k = -1$ for $k \in L(\mathbf{a}, \theta^{\text{old}})$ and $\bar{a}_k = 0$ otherwise, that corresponds to the update rule (51). In this case, the invariant (151) entails (152). The latter is exactly the invariant (53) proven by Tesar and Smolensky for the update rule (51), as reviewed in subsection 4.2.3.1. Thus, here I am really just pointing out that Tesar and Smolensky's invariant does not in any way depend on the fact that the update rule (51) performs demotion only.

Rather, that invariant only depends on the fact that the update rule (51) is minimal (i.e. only demotes currently undominated constraints), and thus essentially carries over to the general minimal update rule (150).

$$(152) \quad \theta_k \geq -k + 1, \quad k = 1, \dots, n$$

Claim 13 does not extend from the specific promotion-demotion update rule (126) to the general promotion-demotion update rule (150). As a counterexample, consider the case of Boersma's promotion-demotion update rule (78), and recall that when that rule is used with Pater's counterexample (82), the components of the current ranking vector keep increasing indefinitely. The crucial property of the update rule (126) that was used in the proof of claim 13 is that the sum over the promotion amounts (namely the amount $\ell(\mathbf{a}, \theta^{\text{old}})$ times the total number $w(\mathbf{a})$ of winner-preferring constraints) is *equal* in size to the sum over the demotion amounts (namely the amount $w(\mathbf{a})$ times the total number $\ell(\mathbf{a}, \theta^{\text{old}})$ of undominated loser-preferring constraints). Of course, the proof of claim 13 given above immediately extends to the case where the sum over the promotion amounts is *not larger* than the sum over the demotion amounts. In other words, the proof extends to the general promotion-demotion update rule (150) provided that the condition (153) holds for every update vector $\bar{\mathbf{a}}$.

$$(153) \quad \sum_{k=1}^n \bar{a}_k \leq 0$$

Claims 14 and 15 extend from the specific promotion-demotion update rule (126) to the general promotion-demotion update rule (150), as long as the update vectors can only take a finite number of integer values. In other words, as long as there exists some bound $B \in \mathbb{N}$ such that condition (154) holds for every update vector $\bar{\mathbf{a}}$.

$$(154) \quad \bar{\mathbf{a}} \in \{-B, \dots, -1, 0, 1, \dots, B\}^n$$

In conclusion, I can generalize claim 15 as claim 16. This conclusion is rather intuitive: promotion-demotion update rules that promote overall less than they demote, in the sense of condition (153), plausibly retain the good convergence properties of demotion-only update rules.

Claim 16 The on-line algorithm (116) run on an input OT-compatible comparative tableau starting from the null initial vector with a promotion-demotion update rule (150) that satisfies the two conditions (153) and (154) can only make a finite number of updates in step 3.

5.3.2 Second variant

An immediate consequence of the general claim 16 is the following claim 17, that guarantees finite time convergence for the case of the promotion-demotion update rules (155) and (156). The former update rule (155) is the update rule already considered in (123), but used here also in the case of comparative rows that have multiple currently undominated loser-preferring constraints; this update rule will be used extensively in chapter 6. The update rule (156) was already considered in Boersma (1998), but the issue of its finite time convergence was open until now. Note that the statement (156) of this update rule does not in any way specify how the unique winner-preferring constraint $C_{\bar{k}}$ that gets promoted should be chosen. We can imagine a variety of ways: we could choose at random; we could stubbornly always choose the same one; we could always choose the constraint corresponding to the smallest index, etcetera. Finite time convergence for all these cases is not intuitively obvious.

(155) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} :

$$\text{a. } \theta_k^{\text{new}} = \begin{cases} \theta_k^{\text{old}} + 1 & \text{if } k \in W(\mathbf{a}) \\ \theta_k^{\text{old}} - w(\mathbf{a}) & \text{if } k \in L(\mathbf{a}, \theta^{\text{old}}) \\ \theta_k^{\text{old}} & \text{otherwise} \end{cases}$$

b. Promote by 1 all winner-preferring constraint; demote all currently undominated loser-preferring constraints by the total number $w(\mathbf{a})$ of winner-preferring constraints.

(156) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} , pick an arbitrary winner-preferring constraint $\bar{k} \in W(\mathbf{a})$ and update as follows:

$$\text{a. } \theta_k^{\text{new}} = \begin{cases} \theta_k^{\text{old}} + 1 & \text{if } k = \bar{k} \in W(\mathbf{a}) \\ \theta_k^{\text{old}} - 1 & \text{if } k \in L(\mathbf{a}, \theta^{\text{old}}) \\ \theta_k^{\text{old}} & \text{otherwise} \end{cases}$$

b. Promote by 1 only the designated winner-preferring constraint $C_{\bar{k}}$; demote by 1 all currently undominated loser-preferring constraints.

Claim 17 The on-line algorithm (116) run on an input OT-compatible comparative tableau starting from the null initial vector with one of the two promotion-demotion update rules (155) or (156) can only perform a finite number of updates in step 3.

Proof. The claim for the case of the update rule (155) is a trivial consequence of the general claim 16, since this update rule (155) is obviously a special case of the general update rule (150) that satisfies the two conditions (153) and (154). The case of the update rule (156) does not strictly speaking constitutes a special case of the general case (150), since the former allows for different updates upon exposure to the same row (depending on the choice of the designated winner-preferring constraint that gets promoted), an option that the scheme (150) does not allow for. But this slight difficulty can be straightforwardly circumvented as follows. Suppose that the input tableau contains a comparative row \mathbf{a} with two entries equal to W corresponding to the two constraints C_h and C_k . Replace that row with two identical copies \mathbf{a}' and \mathbf{a}'' . Define the update vector $\bar{\mathbf{a}}'$ corresponding to the row \mathbf{a}' as in (157a) and the update vector $\bar{\mathbf{a}}''$ corresponding to the row \mathbf{a}'' as in (157b).³

$$(157) \quad \text{a. } \bar{a}'_h = 1 \quad \bar{a}'_k = 0 \quad \bar{a}'_\ell \text{ for every } \ell \in L(\mathbf{a}', \theta^{\text{old}})$$

$$\text{b. } \bar{a}''_h = 0 \quad \bar{a}''_k = 1 \quad \bar{a}''_\ell \text{ for every } \ell \in L(\mathbf{a}'', \theta^{\text{old}})$$

Then, update by row \mathbf{a} and the choice of C_h (respectively, C_k) according to (156) is equivalent to update by row \mathbf{a}' (respectively, \mathbf{a}'') according to (150). ■

5.3.3 Third variant

So far, I have only considered the case of *gradual* update rules. That is indeed the case that I am interested in, because that is the case that is relevant for the OT on-line model of the acquisition of phonology, as discussed in section 4.3.4. Yet, let me close this section with a digression on promotion-demotion *non-gradual* update rules. To simplify the presentation, assume that all rows of the input tableau contain exactly one entry equal to L and let me denote by $C_{L(\mathbf{a})}$ the unique loser-preferring constraint w.r.t. row \mathbf{a} . Claim 2 guarantees that this auxiliary assumption does not affect the generality of the analysis, since a general comparative tableau can be preprocessed and turned into an OT-equivalent comparative tableau with a unique entry equal to L per row. Consider the rule that updates the current ranking vector θ^{old} to the ranking vector θ^{new} in response to a comparative row \mathbf{a} as in (158).

(158) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} :

$$\theta_k^{\text{new}} = \begin{cases} \theta_k^{\text{old}} + \frac{1}{w(\mathbf{a}) + 1} \delta + 1 & \text{if } k \in W(\mathbf{a}) \\ \theta_k^{\text{old}} - \frac{w(\mathbf{a})}{w(\mathbf{a}) + 1} \delta - w(\mathbf{a}) & \text{if } k = L(\mathbf{a}) \\ \theta_k^{\text{old}} & \text{otherwise} \end{cases} \quad \delta = \left[\theta_{L(\mathbf{a})}^{\text{old}} - \max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} \right]$$

This update rule is non-gradual, namely one update by a comparative row \mathbf{a} ensures that the updated vector θ^{new} is right away OT-compatible with \mathbf{a} , as shown by the chain of inequalities in (159).

³Note that it would not do the job to replace the comparative row \mathbf{a} with the two rows \mathbf{a}' and \mathbf{a}'' each of which retains only one of the two w 's, while the other is replaced by an ϵ . In fact, this transformation does not preserve OT-compatibility: the tableau obtained this way could very well not be OT-compatible, even though the original tableau was OT-compatible. If the derived tableau is not OT-compatible, then claim 16 does not apply.

Here, I have reasoned as follows: in step (a), I have used the definition (158) of the update rule; in step (b), I have just shuffled the terms around and simplified; in step (c), I have used the definition of δ .

$$\begin{aligned}
 (159) \quad \max_{h \in W(\mathbf{a})} \theta_h^{\text{new}} - \theta_{L(\mathbf{a})}^{\text{new}} &\stackrel{(a)}{=} \left(\max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} + \frac{1}{w(\mathbf{a}) + 1} \delta + 1 \right) - \left(\theta_{L(\mathbf{a})}^{\text{old}} - \frac{w(\mathbf{a})}{w(\mathbf{a}) + 1} \delta - w(\mathbf{a}) \right) \\
 &\stackrel{(b)}{=} \left(\max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} - \theta_{L(\mathbf{a})}^{\text{old}} \right) + \delta + 1 + w(\mathbf{a}) \\
 &\geq \left| \max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} - \theta_{L(\mathbf{a})}^{\text{old}} \right| + \delta + 1 + w(\mathbf{a}) \\
 &\stackrel{(c)}{=} -\delta + \delta + 1 + w(\mathbf{a}) \\
 &> 0
 \end{aligned}$$

Claim 18 trivially follows from claim 16. The idea is of course to reinterpret an update by the non-gradual update rule (158) as a succession of updates by the gradual rule (123), as in the proof of claim 10 in section 4.2.1.

Claim 18 The on-line algorithm (116) run on an input OT-compatible comparative tableau starting from the null initial vector with the non-gradual promotion-demotion update rules (158) can only perform a finite number of updates in step 3.

Proof. Update triggered by a comparative row \mathbf{a} according to rule (158) can be described as a sequence of $\delta + 1$ updates triggered by that same comparative row according to rule (123), namely (160a) is equivalent to (160b).

$$\begin{array}{c}
 (160) \quad \text{a.} \quad \theta^{\text{old}} \xrightarrow{\mathbf{a}} \theta^{\text{new}} \\
 \quad \quad \quad \downarrow \\
 \quad \quad \quad \text{a and } \theta^{\text{old}} \\
 \quad \quad \quad \text{not OT-comp.} \\
 \quad \quad \quad \downarrow \\
 \quad \quad \quad \text{update} \\
 \quad \quad \quad \text{w.r.t. (158)}
 \end{array}$$

$$\begin{array}{c}
 \text{b.} \quad \theta^{\text{old}} \xrightarrow{\mathbf{a}} \theta^1 \xrightarrow{\mathbf{a}} \theta^2 \xrightarrow{\mathbf{a}} \dots \xrightarrow{\mathbf{a}} \theta^\delta \xrightarrow{\mathbf{a}} \theta^{\text{new}} \\
 \quad \quad \quad \downarrow \quad \quad \quad \downarrow \quad \quad \quad \quad \quad \quad \quad \downarrow \\
 \quad \quad \quad \text{a and } \theta^{\text{old}} \quad \text{a and } \theta^1 \quad \quad \quad \quad \quad \quad \quad \text{a and } \theta^\delta \\
 \quad \quad \quad \text{not OT-comp.} \quad \text{not OT-comp.} \quad \quad \quad \quad \quad \quad \quad \text{not OT-comp.} \\
 \quad \quad \quad \downarrow \quad \quad \quad \downarrow \quad \quad \quad \quad \quad \quad \quad \downarrow \\
 \quad \quad \quad \text{update} \quad \quad \quad \text{update} \quad \quad \quad \quad \quad \quad \quad \text{update} \\
 \quad \quad \quad \text{w.r.t. (123)} \quad \quad \quad \text{w.r.t. (123)} \quad \quad \quad \quad \quad \quad \quad \text{w.r.t. (123)}
 \end{array}$$

Thus, claim 18 immediately follows from claim 15. ■

5.4 Second proof of finite time convergence

In this section, I offer an alternative, very different proof that condition (117) holds for the promotion-demotion update rule (126), namely that the corresponding on-line algorithm (116) can perform only a finite number of updates. This alternative proof is interesting for two reasons. *First*, because it will extend to further promotion-demotion update rules that cannot be analyzed by means of the combinatoric line of proof presented in section 5.2, as discussed in section 5.5. *Second*, because it will shed some light on the interesting issue of the relationship between standard OT and linear OT, as discussed in chapter 8. The reader might indeed want to skip ahead to chapter 8 right after this section. This second line of proof has three steps.

5.4.1 First step

As noted in section 5.1, the new promotion-demotion update rule (126) fits into the additive scheme $\theta^{\text{new}} = \theta^{\text{old}} + \bar{\mathbf{a}}$ in (46) once the update vector $\bar{\mathbf{a}} = (\bar{a}_1, \dots, \bar{a}_n)$, corresponding to a current comparative row \mathbf{a} and a current ranking vector θ^{old} , is defined as in (127), repeated in (161).

$$(161) \quad \bar{a}_k \doteq \begin{cases} \ell(\mathbf{a}, \theta^{\text{old}}) & \text{if } k \in W(\mathbf{a}) \\ -w(\mathbf{a}) & \text{if } k \in L(\mathbf{a}, \theta^{\text{old}}) \\ 0 & \text{otherwise} \end{cases}$$

The core of the second line of analysis of the new promotion-demotion update rule (126) is the following very simple claim 19.

Claim 19 If the ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} , then the inequality (162) holds, where $\bar{\mathbf{a}}$ is the update vector corresponding to θ^{old} and \mathbf{a} according to (161).

$$(162) \quad \langle \theta^{\text{old}}, \bar{\mathbf{a}} \rangle \leq 0$$

In (162), I have used the notation $\langle \cdot, \cdot \rangle$ for the scalar product defined in (94).

Proof. To simplify the notation, let me drop the superscript “old”, and thus write just $\theta = (\theta_1, \dots, \theta_n)$ instead of θ^{old} . The proof of the inequality (162) consists of the chain of inequalities in (163).

$$(163) \quad \begin{aligned} \langle \theta, \bar{\mathbf{a}} \rangle &= \sum_{h=1}^n \theta_h \bar{a}_h \\ &\stackrel{(a)}{=} \sum_{h \in W(\mathbf{a})} \theta_h \bar{a}_h + \sum_{k \in L(\mathbf{a}, \theta)} \theta_k \bar{a}_k + \sum_{h \notin W(\mathbf{a}) \cup L(\mathbf{a}, \theta)} \theta_h \bar{a}_h \\ &\stackrel{(b)}{=} \ell(\mathbf{a}, \theta) \sum_{h \in W(\mathbf{a})} \theta_h - w(\mathbf{a}) \sum_{k \in L(\mathbf{a}, \theta)} \theta_k \\ &\stackrel{(c)}{\leq} \ell(\mathbf{a}, \theta) w(\mathbf{a}) \max_{h \in W(\mathbf{a})} \theta_h - w(\mathbf{a}) \sum_{k \in L(\mathbf{a}, \theta)} \theta_k \\ &\stackrel{(d)}{\leq} \ell(\mathbf{a}, \theta) w(\mathbf{a}) \max_{h \in W(\mathbf{a})} \theta_h - w(\mathbf{a}) \ell(\mathbf{a}, \theta) \min_{k \in L(\mathbf{a}, \theta)} \theta_k \\ &= \ell(\mathbf{a}, \theta) w(\mathbf{a}) \underbrace{\left(\max_{h \in W(\mathbf{a})} \theta_h - \min_{k \in L(\mathbf{a}, \theta)} \theta_k \right)}_{(*)} \\ &\stackrel{(e)}{\leq} 0 \end{aligned}$$

Here, I have reasoned as follows: in step (a), I have split the set $\{1, \dots, n\}$ that h ranges over into the three sets $W(\mathbf{a})$, $L(\mathbf{a}, \theta)$ and their complement; in step (b), I have used the definition (161) of the update vector $\bar{\mathbf{a}}$; in step (c), I have upper bounded the sum $\sum_{h \in W(\mathbf{a})} \theta_h$ with its biggest term $\max_{h \in W(\mathbf{a})} \theta_h$ multiplied by the number $w(\mathbf{a})$ of terms; in step (d), I have lower bounded the sum $\sum_{k \in L(\mathbf{a}, \theta)} \theta_k$ with its smallest term $\min_{k \in L(\mathbf{a}, \theta)} \theta_k$ multiplied by the number $\ell(\mathbf{a}, \theta)$ of terms; in step (e), I have used the fact that θ is by hypothesis not OT-compatible with the comparative row \mathbf{a} , namely there exists a loser-preferring constraint whose ranking value is at least as large as the largest ranking value of winner-preferring constraints, so that $\min_{k \in L(\mathbf{a}, \theta)} \theta_k \geq \max_{h \in W(\mathbf{a})} \theta_h$ and the quantity (*) is therefore nonpositive. ■

5.4.2 Second step

The next claim 20 provides a global properties of the set of all update vectors (161) corresponding to OT-compatible comparative rows. The idea of the proof is rather straightforward. Let me first illustrate it with a concrete example. Consider the OT-compatible comparative tableau \mathbf{A} in (164). For each row \mathbf{a} of this comparative tableau, consider the corresponding update vector $\bar{\mathbf{a}}$ as defined

in (161). Organize these numerical vectors one underneath the other, thus obtaining the numerical matrix $\bar{\mathbf{A}}$ in (161), where I am omitting 0's for the sake of readability.

$$(164) \quad \mathbf{A} = \begin{bmatrix} & C_1 & C_2 & C_3 & C_4 & C_5 \\ W & & & & & \\ & W & & & & \\ & & W & & & \\ & & & W & & L \\ & & & & W & L \\ & & & & & W \\ & & & & & & L \end{bmatrix} \implies \bar{\mathbf{A}} = \begin{bmatrix} & C_1 & C_2 & C_3 & C_4 & C_5 \\ 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & -1 \\ & & & & 1 & -2 \\ & & & & & 1 & -1 \end{bmatrix}$$

Note that the comparative tableau \mathbf{A} is OT-compatible with the ranking $C_1 \gg C_2 \gg \dots \gg C_5$ and is in the form (18). Here is a way to construct a ranking vector θ that satisfies (165). Start from the bottom, and set θ_5 to an arbitrary value, say 0. In order for property (165) to hold for the last row of $\bar{\mathbf{A}}$ with this choice of θ_5 , it must be $\theta_4 \geq 1$; so, let's set $\theta_4 = 1$. In order for property (165) to hold for the third and fourth rows of $\bar{\mathbf{A}}$ with these choices for θ_5 and θ_4 , it must be $\theta_3 \geq 1$; so, let's set $\theta_3 = 1$. And so on. The idea is thus as follows: we arbitrarily pick the ranking value of the bottom ranked constraint and we determine the ranking values of the other constraints one at the time moving upward in the comparative tableau.

Claim 20 If a comparative tableau \mathbf{A} is OT-compatible, then there exists a vector θ such that condition (165) holds for the update vector $\bar{\mathbf{a}}$ corresponding to any row \mathbf{a} of the tableau \mathbf{A} and to any ranking vector according to (161).

$$(165) \quad \langle \theta, \bar{\mathbf{a}} \rangle \geq 1$$

In (165), I have used again the notation $\langle \cdot, \cdot \rangle$ for the scalar product defined in (94).

Proof. Claim 1 ensures that there exists an integer $d \leq n$ such that, by relabeling the constraints and properly reordering the rows and the columns of the comparative tableau \mathbf{A} , it takes the form in (18), repeated once more in (414).

$$(166) \quad \begin{matrix} & C_1 & C_2 & \dots & C_d & C_{d+1} & \dots & C_n \\ dec(C_1) & \left[\begin{array}{cccccccc} W & & & & & & & \\ | & \dots & \dots & \dots & \dots & \dots & \dots & \\ W & & & & & & & \\ E & W & & & & & & \\ | & | & \dots & \dots & \dots & \dots & \dots & \\ E & W & & & & & & \\ E & E & & & & & & \\ \vdots & & & & & & & \\ | & | & \dots & \dots & \dots & \dots & \dots & \\ E & E & & & & & & \\ E & E & & W & & & & \\ | & E & \dots & | & \dots & \dots & \dots & \\ E & E & & W & & & & \end{array} \right] \end{matrix}$$

If the comparative row \mathbf{a} belongs to $dec(C_k)$ for some k , then let me also say that an update vector $\bar{\mathbf{a}}$ corresponding to \mathbf{a} (and to some ranking vector) as in (161) belongs to $dec(C_k)$. Consider the ranking vector $\theta = (\theta_1, \dots, \theta_n)$ defined in (167), starting from the bottom components $\theta_{d+1}, \dots, \theta_n$ and moving up from θ_d to θ_1 . Note that the definition is well-posed because $\bar{a}_k \neq 0$ (in fact, since $\bar{\mathbf{a}} \in dec(C_k)$, then $a_k = w$ and thus $\bar{a}_k > 0$); and furthermore because the maximum is finite (since I am taking the maximum over a finite set).

$$(167) \quad \theta_{d+1} = \dots = \theta_n \doteq 0$$

$$\theta_k \doteq \max_{\bar{\mathbf{a}} \in dec(C_k)} \frac{1}{\bar{a}_k} \left(1 - \sum_{h=k+1}^n \theta_h \bar{a}_h \right), \quad k = d, d-1, \dots, 2, 1$$

Consider an arbitrary update vector $\bar{\mathbf{a}}$ as in (161), and let $k \in \{1, \dots, d\}$ be such that $\bar{\mathbf{a}} \in dec(C_k)$. The chain of implications in (168) shows that condition (165) holds for this update vector $\bar{\mathbf{a}}$.

$$\begin{aligned}
(168) \quad \langle \theta, \bar{a} \rangle \geq 1 &\iff \sum_{h=1}^n \theta_h \bar{a}_h \geq 1 \\
&\stackrel{(a)}{\iff} \sum_{h=1}^{k-1} \theta_h \bar{a}_h + \theta_k \bar{a}_k + \sum_{h=k+1}^n \theta_h \bar{a}_h \geq 1 \\
&\stackrel{(b)}{\iff} \theta_k \bar{a}_k + \sum_{h=k+1}^n \theta_h \bar{a}_h \geq 1 \\
&\iff \theta_k \geq \frac{1}{\bar{a}_k} \left(1 - \sum_{h=k+1}^n \theta_h \bar{a}_h \right) \\
&\stackrel{(c)}{\iff} \theta_k \geq \max_{\bar{a} \in \text{dec}(C_k)} \frac{1}{\bar{a}_k} \left(1 - \sum_{h=k+1}^n \theta_h \bar{a}_h \right)
\end{aligned}$$

Here, I have reasoned as follows: in step (a), I have split up the set $\{1, \dots, n\}$ that h runs over into the three sets $\{1, \dots, k-1\}$, $\{k\}$ and $\{k+1, \dots, n\}$; in step (b), I have used the hypothesis that $\bar{a} \in \text{dec}(C_k)$, that entails that $\bar{a}_1 = \dots = \bar{a}_{k-1} = 0$; in step (c), I have used again the hypothesis that $\bar{a} \in \text{dec}(C_k)$.⁴ ■

5.4.3 Third step

I am now ready to present the following, alternative proof of claim 15 repeated below. The idea of this proof is to deduce finite time convergence of the OT on-line algorithm (116) with the promotion-demotion update rule (126) from the famous convergence theorem for the Perceptron algorithm, recalled as claim 25 in the Appendix that closes this chapter. The specific strategy used here actually illustrates a general strategy to import within standard OT methods and results from the theory of linear classification. This general strategy will be outlined in chapter 8. The interested reader might want to skip ahead to chapter 8 right after this section.

Claim 15 The on-line algorithm (116) run on an input OT-compatible comparative tableau starting from the null initial vector with the promotion-demotion update rule (126) can perform only a finite

⁴Another way of constructing a vector θ that satisfies (165) is (i) below, based on the fact that the components of any update vector (161) cannot be larger (in size) than $n-1$ (in fact, there must be at least one entry equal to L and at least one entry equal to W in every comparative row \mathbf{a} that triggers an update, so that both the number $w(\mathbf{a})$ of winner-preferring constraints and the number $\ell(\mathbf{a}, \theta^{\text{old}})$ of undominated loser-preferring constraints cannot be larger than $n-1$). This observation is in essence due to Prince and Smolensky (2004). Claim 20 is indeed a special version of the well known fact that typologies corresponding to standard OT are smaller than typologies corresponding to linear OT.

$$\begin{aligned}
(i) \quad \theta_1 &= n^n \\
&\vdots \\
\theta_k &= n^{n-k+1} \\
&\vdots \\
\theta_n &= n
\end{aligned}$$

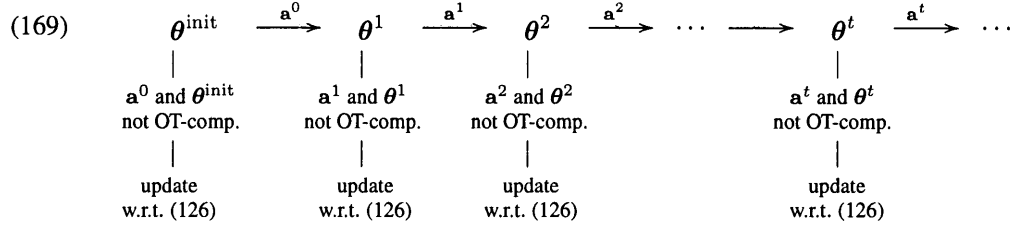
Yet, the ranking vector in (i) has very large components for large n , no matter whether such large components are necessary or not, given a specific comparative tableau. The alternative definition (167) indeed tries to keep the components as small as possible, in the sense that the ranking vector defined in (167) satisfies the component-wise identity in (ii), starting from the bottom component θ_n , and moving up. In fact, since θ_k as defined in (167) belongs to the set $(*)$, then of course $\theta_k \geq \min(*)$. Let $(**)$ be the set obtained by replacing the condition “for every update vector $\bar{\mathbf{a}}$ ” in the definition of $(*)$ with the condition “for every update vector $\bar{\mathbf{a}} \in \text{dec}(C_k)$ ”. Since $(*)$ is a subset of $(**)$, then $\min(**) \leq \min(*)$; furthermore, θ_k as defined in (167) coincides with $\min(**)$; hence $\theta_k \leq \min(*)$.

$$(ii) \quad \theta_k = \min \underbrace{\left\{ \theta'_k \geq 0 \left| \begin{array}{l} \text{there exist } \theta'_1, \dots, \theta'_{k-1} \geq 0 \text{ such that} \\ \theta' \doteq (\theta'_1, \dots, \theta'_{k-1}, \theta'_k, \theta_{k+1}, \dots, \theta_n) \\ \text{satisfies } \langle \theta', \bar{\mathbf{a}} \rangle \geq 1 \text{ for every update vector } \bar{\mathbf{a}} \end{array} \right. \right\}}_{(*)} \quad \text{for } k = n, n-1, \dots, 2, 1$$

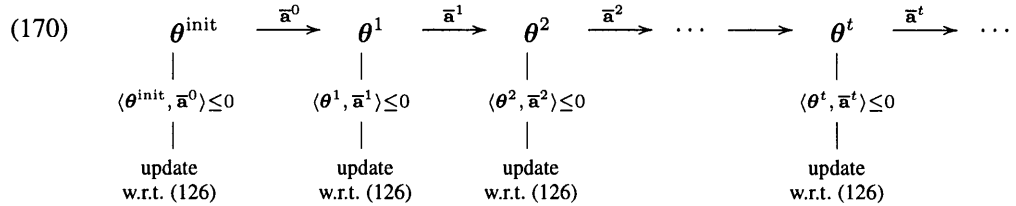
Indeed, it will turn out useful for the developments of subsection 7.1.3, concerning estimates on the actual number of updates, to keep as small as possible the components of the vector θ that ensures (165).

number of updates.

Proof. Assume by contradiction that the claim is false, namely that there exists an input OT-compatible comparative tableau \mathbf{A} with a unique entry equal to L per row such that there exists a way of ordering the rows of this tableau into an infinite sequence $\mathbf{a}^0, \mathbf{a}^1, \dots, \mathbf{a}^t, \dots$ such that, if the rows are fed to the algorithm in this order (namely, the row \mathbf{a}^1 is fed to the algorithm at the first iteration, the row \mathbf{a}^2 at the second iteration, \dots , the row \mathbf{a}^t at the t th iteration, \dots), then at every time t it happens that the current ranking vector θ^t entertained by the algorithm at the t th iteration is not OT-compatible with the comparative row \mathbf{a}^t fed to the algorithm at that iteration, and thus the current ranking vector θ^t gets updated to θ^{t+1} according to the update rule (126). This situation is depicted in (169).



For every time t , let $\bar{\mathbf{a}}^t$ be the update vector corresponding to the current comparative row \mathbf{a}^t and the current ranking vector θ^t as in (161). Since the current ranking vector θ^t is not OT-compatible with the current comparative row \mathbf{a}^t , claim 19 then ensures in particular that $\langle \theta^t, \bar{\mathbf{a}}^t \rangle \leq 0$. The situation depicted in (169) thus in particular entails the situation depicted in (170). The latter situation is entirely described in terms of the update vectors $\bar{\mathbf{a}}^t$: at each time t , we get a vector $\bar{\mathbf{a}}^t$ such that $\langle \theta^t, \bar{\mathbf{a}}^t \rangle \leq 0$ and we update θ^t to θ^{t+1} by adding $\bar{\mathbf{a}}^t$ to θ^t as prescribed by (161).



Claim 20 ensures that the vectors $\bar{\mathbf{a}}^0, \bar{\mathbf{a}}^1, \dots, \bar{\mathbf{a}}^t, \dots$ in (170) satisfy all hypotheses of the convergence theorem of the Perceptron algorithm, namely claim 25 in the Appendix that closes this chapter. Therefore, the latter theorem ensures that the situation depicted in (170) can never arise, thus providing the desired contradiction. ■

5.5 More variants

In this section, I discuss under which conditions the reasoning presented in the preceding section extends from the specific promotion-demotion update rule (126) to a general minimal promotion-demotion update rule of the form (46), repeated in (171). Or equivalently, from the update vector (127) to a general update vector $\bar{\mathbf{a}} = (\bar{a}_1, \dots, \bar{a}_n)$.

(171) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} :

$$\text{a.} \quad \begin{bmatrix} \theta_1^{\text{new}} \\ \vdots \\ \theta_k^{\text{new}} \\ \vdots \\ \theta_n^{\text{new}} \end{bmatrix} = \begin{bmatrix} \theta_1^{\text{old}} \\ \vdots \\ \theta_k^{\text{old}} \\ \vdots \\ \theta_n^{\text{old}} \end{bmatrix} + \begin{bmatrix} \bar{a}_1 \\ \vdots \\ \bar{a}_k \\ \vdots \\ \bar{a}_n \end{bmatrix} \quad \text{where } \bar{a}_k \text{ is } \begin{cases} \leq 0 & \text{if } k \in L(\mathbf{a}, \theta^{\text{old}}) \\ \geq 0 & \text{if } k \in W(\mathbf{a}) \\ = 0 & \text{otherwise} \end{cases}$$

|
update vector

- b. If the constraint C_k is winner-preferring, promote it by adding the nonnegative constant $\bar{a}_k \geq 0$; if the constraint C_k is loser-preferring and undominated, demote it by adding the nonpositive constant $\bar{a}_k \leq 0$.

This discussion will allow me to further comment on and clarify the reasoning presented in the preceding section.

5.5.1 First variant

The proof of claim 15 presented in the preceding section immediately extends to the following generalization.

Claim 21 Consider an update rule of the general form (171) such that there exists a constant $B > 0$ and a ranking vector $\bar{\theta}$ such that for every pair (\mathbf{a}, θ) of a ranking vector θ not OT-compatible with a comparative row \mathbf{a} , the corresponding update vector $\bar{\mathbf{a}}$ satisfies the three properties (172).

- (172) a. $\langle \bar{\mathbf{a}}, \bar{\mathbf{a}} \rangle \leq B$
 b. $\langle \bar{\theta}, \bar{\mathbf{a}} \rangle > 1$
 c. $\langle \theta, \bar{\mathbf{a}} \rangle \leq 0$

The on-line algorithm (116) run on an input OT-compatible comparative tableau starting from the null initial vector with that promotion-demotion update rule can perform only a finite number of updates.

We now need to ensure that the three conditions (172) hold. If the components of the update vectors $\bar{\mathbf{a}} = (\bar{a}_1, \dots, \bar{a}_n)$ can only take a finite number of values, then condition (172a) is trivially satisfied. One way of enforcing that the components of the update vectors can only take a finite number of values is to require them to be bounded integers, as in (173).

$$(173) \quad \bar{\mathbf{a}} \in \{-B, \dots, -1, 0, +1, \dots, B\}^n$$

Consider next the hypothesis (174), that says that for every comparative row $\mathbf{a} = (a_1, \dots, a_n)$ (and for every ranking vector θ), the corresponding update vector $\bar{\mathbf{a}} = (\bar{a}_1, \dots, \bar{a}_n)$ has the property that all the components \bar{a}_k corresponding to winner-preferring constraints are positive. The proof of claim 20 ensures that, if both hypotheses (173) and (174) hold, then condition (172b) holds.

$$(174) \quad \text{If } a_k = w, \text{ then } \bar{a}_k > 0$$

Thus, we only need to worry about condition (172c). Claim 19 showed that condition (172c) does indeed hold in the case of the specific update vectors defined in (161). The proof of claim 19 immediately extends to any definition of the update vectors that satisfies the identity (175a); more in general, it extends to any definition that satisfies the condition (175b).

$$(175) \quad \text{a. } \sum_{h=1}^n \bar{a}_h = 0$$

$$\text{b. } \sum_{h=1}^n \bar{a}_h \begin{cases} \leq 0 & \text{if } \max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} \geq 0 \\ \geq 0 & \text{if } \max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} \leq 0 \end{cases}$$

In conclusion, I can generalize claim 15 as claim 22. Condition (173) is condition (154). And condition (175) is a slight variant of condition (153). Condition (174) is instead peculiar to this line of analysis. In conclusion, claim 22 is a variant of claim 16.

Claim 22 The on-line algorithm (116) run on an input OT-compatible comparative tableau starting from the null initial vector with a promotion-demotion update rule (150) that satisfies the three conditions (173), (174) and (175) can only make a finite number of updates.

5.5.2 Second variant

In subsection 4.2.2, I have reviewed Tesar and Smolensky's observation that finite time convergence (117) holds for minimal update rules with null promotion amounts. In subsection 4.3.4, I have motivated the goal of devising convergent update rules with nonnull promotion amounts. We know that the promotion amounts cannot be too large w.r.t. the demotion amounts, as shown by the failure of Boersma's update rules (78) and (79). So, how large can the promotion amounts be, without affecting convergence? Claim 16 has shown that finite time convergence holds as long as the sum over promotion amounts is not larger than the sum over demotion amounts, as required by condition (153). Claim 22 has further extended this conclusion, by showing that this condition (153) can be relaxed *as long as* the ranking value $\max_{h \in W(\mathbf{a})} \theta_h^{\text{old}}$ of the top ranked winner-preferring constraint is nonpositive, namely that condition (175a) can be replaced by the looser condition (175b). At this point, the following question naturally arises: can we construct update rules where the sum over promotion amounts is larger than the sum over demotion amounts even when the ranking value of the top ranked winner-preferring constraint is nonnegative? Let me rephrase this question more explicitly. Consider the variant (176) of the main update rule (126) studied in this chapter. According to this variant, demotion is only performed when condition (*) holds. In other words, the demotion amounts in cases where the ranking value of the top ranked winner-preferring constraint is nonpositive is as small as possible, namely null. If condition (*) holds and thus demotion is performed, then the demotion-amount is defined as the demotion amount $w(\mathbf{a})$ of the update rule (126) studied so far, multiplied by a constant $\xi_k > 0$. Of course $\xi_k \geq 1$ (respectively, $\xi_k \leq 1$) iff the sum over promotion amounts is smaller (respectively, larger) than the sum over demotion amounts. Thus, we already know that finite time convergence holds for $\xi_k \geq 1$. The question I am interested in can be stated as follows: does there exist a constant $\xi_k \leq 1$ such that finite time convergence holds for the update rule (176)?

(176) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} :

$$\theta_k^{\text{new}} = \begin{cases} \theta_k^{\text{old}} + \ell(\mathbf{a}, \theta^{\text{old}}) & \text{if } k \in W(\mathbf{a}) \\ \theta_k^{\text{old}} - w(\mathbf{a})\xi_k & \text{if } k \in L(\mathbf{a}, \theta^{\text{old}}), \underbrace{\max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} > 0}_{(*)} \\ \theta_k^{\text{old}} & \text{otherwise} \end{cases}$$

Claim 23 offers a positive answer to this question. The constant ξ_k in (177) is well defined and positive. In fact, if demotion is performed, then condition (*) holds, namely there exists at least a winner-preferring constraint with currently positive ranking value; hence, the ranking value θ_k^{old} of the demoted constraint must be positive too, since it must be larger than the ranking value of every winner preferring constraint. Furthermore, $\xi_k \leq 1$, since $\max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} \leq \theta_k$, because $k \in L(\mathbf{a}, \theta^{\text{old}})$. The proof uses the general claim 21.

Claim 23 The on-line algorithm (116) run on an input OT-compatible comparative tableau starting from the null initial vector with the promotion-demotion update rule (176) with the constant ξ defined in (177) can only make a finite number of updates.

$$(177) \quad \xi_k = \frac{\max_{h \in W(\mathbf{a})} \theta_h^{\text{old}}}{\theta_k^{\text{old}}}$$

Proof. To simplify the analysis, divide both the promotion and the demotion amounts of the update rule (176)-(177) by the constant $\ell(\mathbf{a}, \theta^{\text{old}})$. The update rule thus derived fits into the schema (171) with the corresponding update vector $\bar{\mathbf{a}} = (\bar{a}_1, \dots, \bar{a}_n)$ defined in terms of the current ranking vector θ^{old} and the current comparative row \mathbf{a} as in (178).⁵ By claim 21, I just need to show that the three properties (172) hold for the update vectors defined in (178).

⁵The same line of reasoning ensures finite time convergence also for the update rules obtained with one or more of the modifications listed in (i).

- (i) a. Replace the denominator $\ell(\mathbf{a}, \theta^{\text{old}})\theta_k^{\text{old}}$ in (178) with $\sum_{k \in L(\mathbf{a}, \theta^{\text{old}})} \theta_k^{\text{old}}$;
- b. replace $w(\mathbf{a}) \max_{h \in W(\mathbf{a})} \theta_h^{\text{old}}$ in the numerator of (178) with $\sum_{h \in W(\mathbf{a})} \theta_h^{\text{old}}$;

$$(178) \quad \bar{a}_k = \begin{cases} 1 & \text{if } k \in W(\mathbf{a}) \\ -\frac{w(\mathbf{a}) \max_{h \in W(\mathbf{a})} \theta_h^{\text{old}}}{\ell(\mathbf{a}, \theta^{\text{old}}) \theta_k^{\text{old}}} & \text{if } k \in L(\mathbf{a}, \theta^{\text{old}}), \max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} > 0 \\ 0 & \text{otherwise} \end{cases}$$

Let me show that condition (172a) holds. To this end, let me show in particular that $|\bar{a}_k| \leq n$. The only non trivial case is that of the components \bar{a}_k corresponding to currently undominated loser-prefering constraints $k \in L(\mathbf{a}, \theta^{\text{old}})$ under the hypothesis that $\max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} > 0$, since the other components are either 1 or 0 by definition. In this case, the bound is proved by the chain of inequalities in (179). Here, I have reasoned as follows: in step (a), I have used the definition (178); in step (b), I have dropped the absolute value because $\max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} > 0$ by hypothesis and furthermore $\theta_k^{\text{old}} > \max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} > 0$ since $k \in L(\mathbf{a}, \theta^{\text{old}})$; in step (c), I have upper bounded by replacing $\max_{h \in W(\mathbf{a})} \theta_h^{\text{old}}$ by θ_k^{old} , since the latter is larger than the former because $k \in L(\mathbf{a}, \theta^{\text{old}})$.

$$(179) \quad |\bar{a}_k| \stackrel{(a)}{=} \left| \frac{w(\mathbf{a}) \max_{h \in W(\mathbf{a})} \theta_h^{\text{old}}}{\ell(\mathbf{a}, \theta^{\text{old}}) \theta_k^{\text{old}}} \right| \\ \stackrel{(b)}{=} \frac{w(\mathbf{a}) \max_{h \in W(\mathbf{a})} \theta_h^{\text{old}}}{\ell(\mathbf{a}, \theta^{\text{old}}) \theta_k^{\text{old}}} \\ \stackrel{(c)}{\leq} \frac{w(\mathbf{a}) \theta_k^{\text{old}}}{\ell(\mathbf{a}, \theta^{\text{old}}) \theta_k^{\text{old}}} \\ \leq \frac{w(\mathbf{a})}{\ell(\mathbf{a}, \theta^{\text{old}})} \\ \leq n$$

Let me show that condition (172b) holds. Consider the numerical matrix $\bar{\bar{\mathbf{A}}} = [\bar{\bar{a}}_{i,k}] \in \mathbb{R}^{m \times n}$ derived from the input comparative tableau $\mathbf{A} \in \{L, E, W\}^{m \times n}$ as in (180): every entry equal to W in \mathbf{A} is replaced by 1 in $\bar{\bar{\mathbf{A}}}$; every entry equal to E is replaced by 0; and every entry equal to L is replaced by $-n$.

$$(180) \quad \bar{\bar{a}}_{i,k} = \begin{cases} 1 & \text{if } a_{i,k} = W \\ 0 & \text{if } a_{i,k} = E \\ -n & \text{if } a_{i,k} = L \end{cases}$$

By reasoning as in the proof of claim 20, I conclude that there exists a ranking vector $\bar{\theta} = (\bar{\theta}_1, \dots, \bar{\theta}_n)$ such that the strict inequality (181) holds for every row $\bar{\mathbf{a}}$ of the derived numerical matrix $\bar{\bar{\mathbf{A}}}$. Furthermore, the construction used in that proof of claim 20 can straightforwardly be modified to ensure that all components of $\bar{\theta}$ are strictly positive.

$$(181) \quad \langle \bar{\theta}, \bar{\mathbf{a}} \rangle > 0$$

Now let me show that the strict inequality (182) holds for every update vector $\bar{\mathbf{a}}$ defined in (178). Let me distinguish two cases. Let me start with the case where the update vector $\bar{\mathbf{a}}$ corresponds to a current ranking vector θ^{old} and a current comparative row \mathbf{a} such that $\max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} \leq 0$. In this case, all the components of $\bar{\mathbf{a}}$ are either 0 or 1 and at least one of its components is 1 (because each row of the input tableau must admit at least a winner-prefering constraint in order for the tableau to be OT-compatible); since furthermore $\bar{\theta}$ is strictly positive, then the strict inequality (182) holds.

$$(182) \quad \langle \bar{\theta}, \bar{\mathbf{a}} \rangle > 0$$

Let me now turn to the case where the increment vector $\bar{\mathbf{a}}$ corresponds to a current ranking vector θ^{old} and a current comparative row \mathbf{a} such that $\max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} > 0$. In this case, the proof of

c. replace the condition $\max_{h \in W(\mathbf{a})} \theta_h^{\text{old}} > 0$ in (*) with the condition $\sum_{h \in W(\mathbf{a})} \theta_h^{\text{old}} > 0$.

the inequality (182) consists of the chain of inequalities in (183). Here, I have reasoned as follows: in step (a), I have used the definition (94) of scalar product; in step (b), I have split up the set $\{1, \dots, n\}$ that h runs over into the three sets $W(\mathbf{a})$, $L(\mathbf{a}, \theta^{\text{old}})$ and their complement; in step (c), I have used the definition (178), that ensures that $\bar{a}_h = 1$ for every $h \in W(\mathbf{a})$ and $\bar{a}_h = 0$ for every $h \notin W(\mathbf{a}) \cup L(\mathbf{a}, \theta^{\text{old}})$; in step (d), I have used the fact that $\bar{a}_h \geq -n$ for every $h \in L(\mathbf{a}, \theta^{\text{old}})$, as proven in (179); in step (e), I have used the fact that $L(\mathbf{a}, \theta^{\text{old}}) \subseteq L(\mathbf{a})$; in step (f), I have used the definition (180) of the numerical vector $\bar{\mathbf{a}}$ corresponding to the comparative row \mathbf{a} ; in the last step (g), I have used the hypothesis (181).

$$\begin{aligned}
 (183) \quad \langle \bar{\theta}, \bar{\mathbf{a}} \rangle &\stackrel{(a)}{=} \sum_{h=1}^n \bar{\theta}_h \bar{a}_h \\
 &\stackrel{(b)}{=} \sum_{h \in W(\mathbf{a})} \bar{\theta}_h \bar{a}_h + \sum_{h \in L(\mathbf{a}, \theta^{\text{old}})} \bar{\theta}_h \bar{a}_h + \sum_{h \notin W(\mathbf{a}) \cup L(\mathbf{a}, \theta^{\text{old}})} \bar{\theta}_h \bar{a}_h \\
 &\stackrel{(c)}{=} \sum_{h \in W(\mathbf{a})} \bar{\theta}_h + \sum_{h \in L(\mathbf{a}, \theta^{\text{old}})} \bar{\theta}_h \bar{a}_h \\
 &\stackrel{(d)}{\geq} \sum_{h \in W(\mathbf{a})} \bar{\theta}_h + \sum_{h \in L(\mathbf{a}, \theta^{\text{old}})} \bar{\theta}_h (-n) \\
 &\stackrel{(e)}{\geq} \sum_{h \in W(\mathbf{a})} \bar{\theta}_h + \sum_{h \in L(\mathbf{a})} \bar{\theta}_h (-n) \\
 &\stackrel{(f)}{\geq} \langle \bar{\theta}, \bar{\mathbf{a}} \rangle \\
 &\stackrel{(g)}{>} 0
 \end{aligned}$$

Let me show that condition (172c) holds. Consider a current comparative row \mathbf{a} and a current ranking vector θ^{old} not OT-compatible with it; let $\bar{\mathbf{a}}$ be the corresponding update vector (178). The proof of condition (172c) consists of the chain of inequalities in (184). Here, I have reasoned as follows: in step (a), I have used the definition (94) of scalar product; in step (b), I have split up the set $\{1, \dots, n\}$ that h runs over into the three sets $W(\mathbf{a})$, $L(\mathbf{a}, \theta^{\text{old}})$ and their complement; in step (c), I have used the definition (178); in step (d), I have used the fact that $\ell(\mathbf{a}, \theta^{\text{old}})$ is the cardinality of the set $L(\mathbf{a}, \theta^{\text{old}})$.

$$\begin{aligned}
 (184) \quad \langle \theta^{\text{old}}, \bar{\mathbf{a}} \rangle &\stackrel{(a)}{=} \sum_{h=1}^n \theta_h^{\text{old}} \bar{a}_h \\
 &\stackrel{(b)}{=} \sum_{h \in W(\mathbf{a})} \theta_h^{\text{old}} \bar{a}_h + \sum_{h \in L(\mathbf{a}, \theta^{\text{old}})} \theta_h^{\text{old}} \bar{a}_h + \sum_{h \notin W(\mathbf{a}) \cup L(\mathbf{a}, \theta^{\text{old}})} \theta_h^{\text{old}} \bar{a}_h \\
 &\stackrel{(c)}{=} \sum_{h \in W(\mathbf{a})} \theta_h^{\text{old}} - \sum_{h \in L(\mathbf{a}, \theta^{\text{old}})} \theta_h^{\text{old}} \left(\frac{w(\mathbf{a}) \max_{k \in W(\mathbf{a})} \theta_k^{\text{old}}}{\ell(\mathbf{a}, \theta^{\text{old}}) \theta_h^{\text{old}}} \right) \\
 &= \sum_{h \in W(\mathbf{a})} \theta_h^{\text{old}} - w(\mathbf{a}) \max_{k \in W(\mathbf{a})} \theta_k^{\text{old}} \sum_{h \in L(\mathbf{a}, \theta^{\text{old}})} \theta_h^{\text{old}} \left(\frac{1}{\ell(\mathbf{a}, \theta^{\text{old}}) \theta_h^{\text{old}}} \right) \\
 &\stackrel{(d)}{=} \sum_{h \in W(\mathbf{a})} \theta_h^{\text{old}} - w(\mathbf{a}) \max_{k \in W(\mathbf{a})} \theta_k^{\text{old}} \\
 &\leq 0
 \end{aligned}$$

Note the following difference between the two chain inequalities in (163) and (184) used to prove that $\langle \theta^{\text{old}}, \bar{\mathbf{a}} \rangle \leq 0$ in the case of the two update rules (126) and (176)-(177), respectively: in the former case, I used the hypothesis that θ^{old} is not OT-compatible with \mathbf{a} ; in the latter case, I did not use that hypothesis, namely the inequality holds no matter what. ■

5.5.3 Third variant

So far, I have considered update rules that fit into the additive scheme (46). Let me close this section by considering an update rule (42) for the OT on-line algorithm that does not fall into this broad schema. Given a comparative tableau \mathbf{A} , pair up each row \mathbf{a} of the tableau with a numerical vector $\bar{\mathbf{a}}$. Consider the update rule (185), defined in terms of these update vectors $\bar{\mathbf{a}}$.

(185) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} :

$$\theta_k^{\text{new}} = \frac{1}{Z} \theta_k^{\text{old}} \exp \left\{ \frac{\epsilon}{D(\bar{\mathbf{a}})} \bar{a}_k \right\}$$

where I have used the following positions:

- a. Z is the normalization coefficient;
- b. $\epsilon > 0$ is a positive constant;
- c. $D(\bar{\mathbf{a}}) = \max_{h \in W(\mathbf{a})} \bar{a}_h + \max_{h \in L(\mathbf{a})} (-\bar{a}_h)$ is the sum of the maximum promotion amount and the maximum demotion amount.

Assume that the update vector $\bar{\mathbf{a}} = (\bar{a}_1, \dots, \bar{a}_n) \in \mathbb{R}^n$ corresponding to a comparative row $\mathbf{a} = (a_1, \dots, a_n) \in \{L, E, W\}^n$ satisfies the following condition (186), that in particular entails condition (174). This condition (186) ensures that $D(\bar{\mathbf{a}})$ is never null, so that the definition (185) makes sense.

(186) For $k = 1, \dots, n$:

- a. if $a_k = W$, then $\bar{a}_k > 0$;
- b. if $a_k = E$, then $\bar{a}_k = 0$;
- c. if $a_k = L$, then $\bar{a}_k \leq 0$.

The current ranking vector entertained by the on-line algorithm (116) with the update rule (185) is nonnegative by construction. Thus, condition (175) in this case boils down to (187).

$$(187) \quad \sum_{h=1}^n \bar{a}_h \leq 0$$

Choose the constant ϵ in (185) in such a way that there exists a probability vector θ such that the inequality (188) holds for every update vector $\bar{\mathbf{a}}$. Such a constant can be determined by ℓ_1 -normalizing the nonnegative vector constructed in (167).

$$(188) \quad \langle \theta, \bar{\mathbf{a}} \rangle \geq \epsilon$$

The proof of claim 24 is identical to the proof of claim 22. The only difference is that it uses the linear on-line algorithm described by claim 26 in the Appendix that closes this chapter, instead of the Perceptron algorithm described in claim 25. I provide it explicitly for completeness. In this work, I do not study this update rule (185). See Magri (2009) for a small application of this update rule to modeling the data in (101).

Claim 24 The on-line algorithm (116) run on an input OT-compatible comparative tableau starting from the initial vector $\theta^{\text{init}} = (\frac{1}{n}, \dots, \frac{1}{n})$ with the update rule (185) that satisfies the four conditions (173), (186), (187) and (188) can only make a finite number of updates.

Proof. Assume by contradiction that the claim is false, namely that there exists an input OT-compatible comparative tableau \mathbf{A} such that there exists a way of ordering the rows of this tableaux into an infinite sequence $\mathbf{a}^0, \mathbf{a}^1, \dots, \mathbf{a}^t, \dots$ such that, if the rows are fed to the algorithm in this order (namely, the row \mathbf{a}^1 is fed to the algorithm at the first iteration, the row \mathbf{a}^2 at the second iteration, ..., the row \mathbf{a}^t at the t th iteration, ...), then at every time t it happens that the current ranking vector θ^t entertained by the algorithm at the t th iteration is not OT-compatible with the comparative row \mathbf{a}^t fed to the algorithm at that iteration, and thus the current ranking vector θ^t gets updated to θ^{t+1} according to the update rule (185). Let $\bar{\mathbf{a}}^t$ be the update vector used at time t . This situation is depicted in (189).

$$(189) \quad \begin{array}{ccccccc} \theta^{\text{init}} & \xrightarrow{\bar{a}^0} & \theta^1 & \xrightarrow{\bar{a}^1} & \theta^2 & \xrightarrow{\bar{a}^2} & \dots \longrightarrow \theta^t \xrightarrow{\bar{a}^t} \dots \\ | & & | & & | & & | \\ \mathbf{a}^0 \text{ and } \theta^{\text{init}} & & \mathbf{a}^1 \text{ and } \theta^1 & & \mathbf{a}^2 \text{ and } \theta^2 & & \mathbf{a}^t \text{ and } \theta^t \\ \text{not OT-comp.} & & \text{not OT-comp.} & & \text{not OT-comp.} & & \text{not OT-comp.} \\ | & & | & & | & & | \\ \text{update} & & \text{update} & & \text{update} & & \text{update} \\ \text{w.r.t. (185)} & & \text{w.r.t. (185)} & & \text{w.r.t. (185)} & & \text{w.r.t. (185)} \end{array}$$

At each time t the current ranking vector θ^t is not OT-compatible with the current comparative row \mathbf{a}^t . The hypothesis (187) together with the fact that the current ranking vector θ^t is nonnegative by construction ensure that $\langle \theta^t, \bar{\mathbf{a}}^t \rangle \leq 0$ for every vector in the infinite sequence $\bar{\mathbf{a}}^0, \bar{\mathbf{a}}^1, \dots, \bar{\mathbf{a}}^t, \dots$. The proof of this fact is a trivial variant of the chain of inequalities in (163). The situation in (189) thus entails the situation in (190).

$$(190) \quad \begin{array}{ccccccc} \theta^{\text{init}} & \xrightarrow{\bar{a}^0} & \theta^1 & \xrightarrow{\bar{a}^1} & \theta^2 & \xrightarrow{\bar{a}^2} & \dots \longrightarrow \theta^t \xrightarrow{\bar{a}^t} \dots \\ | & & | & & | & & | \\ \langle \theta^{\text{init}}, \bar{\mathbf{a}}^0 \rangle \leq 0 & & \langle \theta^1, \bar{\mathbf{a}}^1 \rangle \leq 0 & & \langle \theta^2, \bar{\mathbf{a}}^2 \rangle \leq 0 & & \langle \theta^t, \bar{\mathbf{a}}^t \rangle \leq 0 \\ | & & | & & | & & | \\ \text{update} & & \text{update} & & \text{update} & & \text{update} \\ \text{w.r.t. (185)} & & \text{w.r.t. (185)} & & \text{w.r.t. (185)} & & \text{w.r.t. (185)} \end{array}$$

The two hypotheses (173 and (188), together with the hypothesis that the input comparative tableau is OT-compatible, ensure that the hypotheses of claim 26 in the Appendix that closes this chapter are all satisfied. The latter claim says that the situation depicted in (190) can never arise, thus providing the desired contradiction. ■

Appendix: some classical results on linear on-line algorithms

The proof of claim 15 provided at the end of section 5.4 rests on the the following classical claim 25, also known as the convergence theorem for the perceptron algorithm in the separable case. I recall the proof for completeness; see for instance Cristianini and Shawe-Taylor (2000, Theorem 2.3).

Claim 25 Given a sequence $\{\bar{\mathbf{a}}^t\}_{t=0}^{\infty}$ of vectors $\bar{\mathbf{a}}^t = (\bar{a}_1^t, \dots, \bar{a}_n^t) \in \mathbb{R}^n$, consider the corresponding sequence $\{\theta^t\}_{t=1}^{\infty}$ of vectors $\theta^t = (\theta_1^t, \dots, \theta_n^t) \in \mathbb{R}^n$ defined as in (191).

$$(191) \quad \begin{aligned} \theta^0 &= \mathbf{0} \\ \theta^{t+1} &= \theta^t + \bar{\mathbf{a}}^t \end{aligned}$$

There exists no infinite sequence $\{\bar{\mathbf{a}}^t\}_{t=0}^{\infty} \subseteq \mathbb{R}^n$ such that the the three properties (192) hold.

- $$(192) \quad \begin{aligned} \text{a.} & \text{ there exists a vector } \theta = (\theta_1, \dots, \theta_n) \in \mathbb{R}^n \text{ and a positive constant } \mu > 0 \text{ such that} \\ & \langle \theta, \bar{\mathbf{a}}_t \rangle / \|\theta\| \geq \mu \text{ for every } t = 1, 2, \dots; \\ \text{b.} & \text{ there exists a constant } R \text{ such that } \|\bar{\mathbf{a}}_t\|^2 \leq R^2 \text{ for every } t = 1, 2, \dots; \\ \text{c.} & \langle \theta^t, \bar{\mathbf{a}}_t \rangle \leq 0 \text{ for every } t = 1, 2, \dots \end{aligned}$$

Proof. The proof has three parts. The *first* part of the proof estimates the norm of θ^t as in (193). Here, I have reasoned as follows: in step (a), I have used the definition (191); in step (b), I have used the identity $\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 + 2\langle \mathbf{v}, \mathbf{w} \rangle$; in step (c), I have used the hypothesis (192b); in step (d), I have used the hypothesis (192c). The inequality (193b) immediately follows from (193a) together with the hypothesis that $\theta^0 = \mathbf{0}$.

$$(193) \quad \begin{aligned} \text{a.} \quad \|\theta^{t+1}\|^2 &\stackrel{(a)}{=} \|\theta^t + \bar{\mathbf{a}}_{i_t}\|^2 \\ &\stackrel{(b)}{=} \|\theta^t\|^2 + \|\bar{\mathbf{a}}_{i_t}\|^2 + 2\langle \theta^t, \bar{\mathbf{a}}_{i_t} \rangle \\ &\stackrel{(c)}{\leq} \|\theta^t\|^2 + R^2 + 2\langle \theta^t, \bar{\mathbf{a}}_{i_t} \rangle \\ &\stackrel{(d)}{\leq} \|\theta^t\|^2 + R^2 \end{aligned}$$

$$\text{b. } \|\theta^t\|^2 \leq tR^2$$

The *second* part of the proof estimates the scalar product $\langle \theta, \theta^{t+1} \rangle$ as in (194). Here, I have reasoned as follows: in step (a), I have used the definition (191); in step (b), I have used the linearity of the scalar product; in step (c), I have used the hypothesis (192c). The inequality (194b) immediately follows from (194a) together with the hypothesis that $\theta^0 = 0$.

$$(194) \quad \text{a. } \frac{\langle \theta, \theta^{t+1} \rangle}{\|\theta\|} \stackrel{(a)}{=} \frac{\langle \theta, \theta^t + \bar{a}_{i_t} \rangle}{\|\theta\|} \\ \stackrel{(b)}{=} \frac{\langle \theta, \theta^t \rangle}{\|\theta\|} + \frac{\langle \theta, \bar{a}_{i_t} \rangle}{\|\theta\|} \\ \stackrel{(c)}{\geq} \frac{\langle \theta, \theta^t \rangle}{\|\theta\|} + \mu \\ \text{b. } \frac{\langle \theta, \theta^t \rangle}{\|\theta\|} \geq t\mu.$$

The *third* part of the proof connects the two inequalities (193b) and (194b) as in (195). Here, I have reasoned as follows: in step (a), I have used (194b); in step (b), I have used the Cauchy-Schwartz Inequality; in step (c), I have used (193b).

$$(195) \quad (t\mu)^2 \stackrel{(a)}{\leq} \left(\frac{\langle \theta, \theta^t \rangle}{\|\theta\|} \right)^2 \\ \stackrel{(b)}{\leq} \frac{\|\theta\|^2 \|\theta^t\|^2}{\|\theta\|^2} \\ = \|\theta^t\|^2 \\ \stackrel{(c)}{\leq} tR^2$$

The chain of inequalities (195) entails in particular that $t \leq R^2/\mu^2$. ■

The proof of claim 24 rests on the the following classical claim 26, fully parallel to the preceding claim 25. Here, I report the proof of claim 26 provided in Cesa-Bianchi (1997, Lemma 3). See Cesa-Bianchi and Lugosi (2006, Ch. 12) for a broader perspective.

Claim 26 Given a sequence $\{\bar{a}^t\}_{t=0}^\infty$ of vectors $\bar{a}^t = (\bar{a}_1^t, \dots, \bar{a}_n^t) \in \mathbb{R}^n$, consider the corresponding sequence $\{\theta^t\}_{t=1}^\infty$ of probability vectors $\theta^t = (\theta_1^t, \dots, \theta_n^t) \in \Delta_n$ defined as in (196), where Z is the corresponding normalization coefficient and $\epsilon, R > 0$ are positive constants.

$$(196) \quad \theta_k^0 = \frac{1}{n} \\ \theta_k^{t+1} = \frac{1}{Z} \theta_k^t \exp \left\{ \frac{\epsilon}{R^2} \bar{a}_k^t \right\}$$

There exists no infinite sequence $\{\bar{a}^t\}_{t=0}^\infty \subseteq \mathbb{R}^n$ such that the the three properties (197) hold.

$$(197) \quad \text{a. There exists a probability vector } \bar{\theta} \text{ such that } \langle \bar{\theta}, \bar{a}^t \rangle \geq \epsilon \text{ for every } t = 1, 2, \dots; \\ \text{b. } D(\mathbf{a}^t) = \max_{h=1, \dots, n} \bar{a}_h^t - \min_{h=1, \dots, n} \bar{a}_h^t \leq R^2 \text{ for every } t = 1, 2, \dots; \\ \text{c. } \langle \theta^t, \bar{a}^t \rangle \leq 0 \text{ for every } t = 1, 2, \dots$$

First part of the proof. The crucial ingredient of the proof is the inequality (198), that holds for any probability vector $\theta \in \Delta_n$. The quantity $D(\cdot|\cdot)$ is the Kullback-Liebler divergence, defined as $D(\mathbf{p}|\mathbf{q}) \doteq \sum_{i=1}^n p_i \log \frac{p_i}{q_i}$ for any two probability vectors $\mathbf{p}, \mathbf{q} \in \Delta_n$. I will review the proof of (198) in the second part of the proof.

$$(198) \quad \langle \theta - \theta^t, \bar{a}^t \rangle \leq \frac{R^2}{\epsilon} \left(D(\theta|\theta^t) - D(\theta|\theta^{t+1}) \right) + \frac{\epsilon}{8}$$

The chain of inequalities in (199) holds for every $T = 1, 2, \dots$. Here, I have reasoned as follows: in step (a), I have used the hypothesis (197a) that $\langle \bar{\theta}, \bar{\mathbf{a}}^t \rangle \geq \epsilon$, that entails that $\frac{1}{\epsilon} \langle \bar{\theta}, \bar{\mathbf{a}}^t \rangle \geq 1$; in step (b), I have used the hypothesis (197c) that $\langle \theta^t, \bar{\mathbf{a}}^t \rangle \leq 0$; in step (c), I have used (198); in step (d), I have used the fact that the sum (*) is telescopic; in the step (e), I have used the fact that $D(\cdot|\cdot) \geq 0$.

$$\begin{aligned}
(199) \quad T &\stackrel{(a)}{\leq} \frac{1}{\epsilon} \sum_{t=1}^T \langle \bar{\theta}, \bar{\mathbf{a}}^t \rangle \\
&\stackrel{(b)}{\leq} \frac{1}{\epsilon} \sum_{t=1}^T \left(\langle \bar{\theta}, \bar{\mathbf{a}}^t \rangle - \langle \theta^t, \bar{\mathbf{a}}^t \rangle \right) \\
&= \frac{1}{\epsilon} \sum_{t=1}^T \langle \bar{\theta} - \theta^t, \bar{\mathbf{a}}^t \rangle \\
&\stackrel{(c)}{\leq} \underbrace{\frac{R^2}{\epsilon^2} \sum_{t=1}^T \left(D(\theta|\theta^t) - D(\theta|\theta^{t+1}) \right)}_{(*)} + \sum_{t=1}^T \frac{1}{8} \\
&\stackrel{(d)}{=} \frac{R^2}{\epsilon^2} \left(D(\theta|\theta^1) - D(\theta|\theta^{T+1}) \right) + \frac{1}{8}T \\
&\stackrel{(e)}{\leq} \frac{R^2}{\epsilon^2} D(\theta|\theta^1) + \frac{1}{8}T
\end{aligned}$$

The inequality derived through (199) entails that T must be finite.

Second part of the proof. Let me now turn to the proof of the crucial inequality (198). The bulk of the proof of (198) consists of the chain of inequalities in (200). Here, I have reasoned as follows: in step (a), I have used the definition (196) of θ_k^{t+1} , with the position $\eta = \epsilon/R^2$ to simplify notation; in step (b), I have used the fact that the $\sum_{i=1}^n \theta_i = 1$; in step (c), I have made explicit the normalization coefficient Z ; in step (d), I have added and subtracted the quantity $\eta \langle \theta^t, \bar{\mathbf{a}}^t \rangle$ from the exponent.

$$\begin{aligned}
(200) \quad D(\theta|\theta^t) - D(\theta|\theta^{t+1}) &= \\
&= \sum_{h=1}^n \theta_h \log \frac{\theta_h}{\theta_h^t} - \sum_{i=1}^n \theta_h \log \frac{\theta_h}{\theta_h^{t+1}} \\
&= \sum_{h=1}^n \theta_h \log \frac{\theta_h^{t+1}}{\theta_h^t} \\
&\stackrel{(a)}{=} \sum_{h=1}^n \theta_h \log \frac{\frac{1}{Z} \theta_h^t e^{\eta \bar{\mathbf{a}}_h^t}}{\theta_h^t} \\
&= \eta \sum_{h=1}^n \theta_h \bar{\mathbf{a}}_h^t - \sum_{i=1}^n \theta_h \log Z \\
&\stackrel{(b)}{=} \eta \langle \theta, \bar{\mathbf{a}}^t \rangle - \log Z \\
&\stackrel{(c)}{=} \eta \langle \theta, \bar{\mathbf{a}}^t \rangle - \log \left(\sum_{h=1}^n \theta_h^t e^{\eta \bar{\mathbf{a}}_h^t} \right) \\
&\stackrel{(d)}{=} \eta \langle \theta, \bar{\mathbf{a}}^t \rangle - \log \left(\sum_{h=1}^n \theta_h^t e^{\eta \bar{\mathbf{a}}_h^t + \eta \langle \theta^t, \bar{\mathbf{a}}^t \rangle - \eta \langle \theta^t, \bar{\mathbf{a}}^t \rangle} \right) \\
&= \eta \langle \theta, \bar{\mathbf{a}}^t \rangle - \log \left(e^{\eta \langle \theta^t, \bar{\mathbf{a}}^t \rangle} \sum_{h=1}^n \theta_h^t e^{\eta \bar{\mathbf{a}}_h^t - \eta \langle \theta^t, \bar{\mathbf{a}}^t \rangle} \right) \\
&= \eta \langle \theta, \bar{\mathbf{a}}^t \rangle - \eta \langle \theta^t, \bar{\mathbf{a}}^t \rangle - \underbrace{\log \left(\sum_{h=1}^n \theta_h^t e^{\eta (\bar{\mathbf{a}}_h^t - \langle \theta^t, \bar{\mathbf{a}}^t \rangle)} \right)}_{(*)}
\end{aligned}$$

To conclude the proof, I need to bound the term (*) in (200). To this end, we use the following fact: if Y is a random variable with zero mean and range $[a, b]$, then the inequality (201) holds for every scalar $t \in \mathbb{R}$; see Cesa-Bianchi (1997, Lemma 2).

$$(201) \quad \log \mathbb{E}[e^{tY}] \leq \frac{t^2}{8}(b-a)^2$$

Consider the random variable Y defined as follows: it takes values $\bar{a}_h^t - \langle \theta^t, \bar{a}^t \rangle$ for $h = 1, \dots, n$ with probability θ_h^t . The expected value of the random variable Y is null, as shown in (202), where in step (a) I have used the fact that θ^t is a probability vector.

$$\begin{aligned} (202) \quad \mathbb{E}[Y] &= \sum_{h=1}^n \theta_h^t (\bar{a}_h^t - \langle \theta^t, \bar{a}^t \rangle) \\ &= \sum_{h=1}^n \theta_h^t \bar{a}_h^t - \sum_{i=1}^n \theta_i^t \langle \theta^t, \bar{a}^t \rangle \\ &\stackrel{(a)}{=} \langle \theta^t, \bar{x}^t \rangle - \langle \theta^t, \bar{a}^t \rangle \\ &= 0 \end{aligned}$$

And the range $[a, b]$ of Y can be described as in (203).

$$\begin{aligned} (203) \quad b - a &= \max_{h=1, \dots, n} (\bar{a}_h^t - \langle \theta^t, \bar{a}^t \rangle) - \min_{h=1, \dots, n} (\bar{a}_h^t - \langle \theta^t, \bar{a}^t \rangle) \\ &= \left(\max_{h=1, \dots, n} \bar{a}_h^t \right) - \left(\min_{h=1, \dots, n} \bar{a}_h^t \right) \\ &= D(\bar{a}^t) \end{aligned}$$

I can thus bound (*) by means of the chain of inequalities in (204). Here, I have reasoned as follows: in step (a), I have introduced the random variable Y just defined; in step (b), I have used (201); in step (c), I have used the hypothesis (197b).

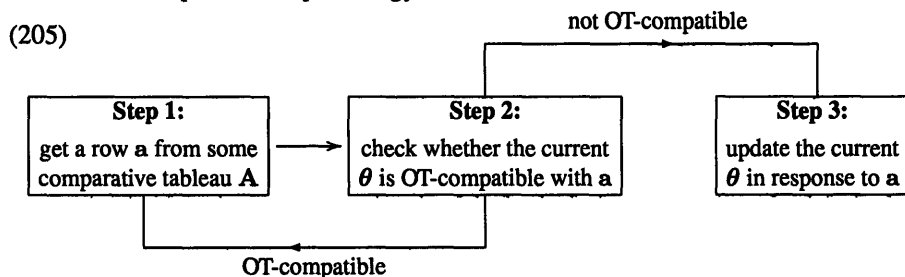
$$\begin{aligned} (204) \quad \log \left(\sum_{i=1}^n \theta_i^t e^{\eta(\bar{a}_i^t - \langle \theta^t, \bar{a}^t \rangle)} \right) &\stackrel{(a)}{=} \log \mathbb{E}[e^{\eta Y}] \\ &\stackrel{(b)}{\leq} \frac{\eta^2}{8} D(\bar{a}^t)^2 \\ &\stackrel{(c)}{\leq} \frac{\epsilon^2}{8R^4} R^2 \\ &= \end{aligned}$$

The inequality (198) thus follows by combining the two inequalities (200) and (204). ■

Chapter 6

How to study constraint promotion: the final ranking vector

Chapter 4 has introduced Hayes' (2004) problem of modeling the early stage of the acquisition of phonology and has suggested that, in order for the OT on-line algorithm (205) to count as a suitable model, it needs update rules in step 3 that perform promotion too. The goal of this second part of the dissertation is to develop the beginning of a theory of constraint promotion for the OT on-line model of the acquisition of phonology.



As noted in subsection 4.2.1, a theory of a given update rule for step 3 of the OT on-line algorithm (205) should address (at least) the three questions (45), repeated in (206). Chapter 5 has introduced a number of new promotion-demotion update rules and has presented different techniques to show that they can only trigger a finite number of updates, thus settling the issue (206a). This chapter and the next one start the research project of studying the properties of these promotion-demotion update rules, by focusing on the two issues (206b) and (206c), respectively.

- (206)
- Does the OT on-line algorithm (205) with a given update rule converge in finite time, namely is true that it can only perform a finite number of updates when run on an OT-compatible input comparative tableau?
 - If it does converge, how can the final vector θ^{fin} be characterized in terms of the input comparative tableau A and the initial ranking vector θ^{init} ?
 - And what is the worst case number of updates in step 3 for an arbitrary input comparative tableau A with n columns and an arbitrary initial vector θ^{init} ?

In this chapter, I concentrate on the issue (206b) concerning the characterization of the final vector entertained by the OT on-line algorithm (205) run with a promotion-demotion update rule in step 3. As reviewed in subsection 4.2.3.2, the issue (206b) of the characterization of the final vector was easy to settle in the case of (minimal) demotion-only update rules. The strategy used in that case was to look for (PRIMAL) INVARIANTS, namely for properties that hold of the current ranking vector θ^t entertained by the algorithm (205) at any time t . If a property holds of the current ranking vector at any time, then it holds of the final vector θ^{fin} . In the case of minimal demotion-only update rules, this simple strategy led to the very sharp characterization of the final vector in claim 8, as the component-wise maximum over all integral non-positive ranking vectors OT-compatible with the input tableau

A. In section 6.1, I try to replicate this strategy for promotion-demotion update rules. I present a simple DUAL INVARIANT, namely a property that holds of the current dual vector α^t entertained at a generic time t by the algorithm (205) run with promotion-demotion update rules. Again, this dual invariant applies in particular to the final dual vector α^{fin} . And the final dual vector α^{fin} can in turn be used to characterize the final primal vector θ^{fin} , since the dual vector always determines the primal vector, as noted in subsection 4.2.6. Unfortunately, contrary to the case of demotion-only update rules, the characterization obtained in this way for the case of promotion-demotion update rules turns out to be rather loose in the general case. In section 6.3, I thus switch gears: instead of tackling question (206b) in the general case, I turn to a few specific case studies. In particular, I study the final vector entertained by the on-line algorithm (205) with promotion-demotion update rules in the case of input phonotactics comparative tableaux, as defined in section 6.2. I consider both artificial phonotactics tableau and “naturalistic” ones. The discussion of these case studies has two goals. *First*, to illustrate various methods to study the behavior of the algorithm in full detail with paper and pencil, by exploiting in various ways the invariant introduced in section 6.1. *Second*, to put forward a bold but very preliminary conjecture, namely that the strong symmetry of the W 's in naturalistic phonotactics tableaux, coupled with the sensitivity of promotion-demotion update rules to those W 's, might actually help the algorithm to reach the correct final ranking vector.

6.1 An invariant

To simplify the discussion, I will assume throughout this section that the input comparative tableau $\mathbf{A} \in \{L, E, W\}^{m \times n}$ contains exactly one entry equal to L per row. Claim 2 guarantees that this auxiliary assumption does not affect the generality of the analysis, since a general comparative tableau can be preprocessed and turned into an OT-equivalent comparative tableau with a unique entry equal to L per row. For concreteness, I will concentrate on the promotion-demotion update rule introduced in section 5.1, that in this case takes the form (123), repeated in (207); yet, the reasoning presented in this section extends to any promotion-demotion update rule. For the case of (minimal) demotion-only update rules, Tesar and Smolensky were able to derive the primal invariants stated in claims 5 and 7. These invariants say that the components of the current ranking vector cannot become too large (in absolute value) over time. As noted in subsection 4.2.6, once we switch from demotion-only to promotion-demotion update rules, it is easier to carry out theoretical analyses at the dual level rather than at the primal level. In this section, I thus look for dual invariants for the on-line algorithm (205) with the promotion-demotion update rule (207). Again, the idea is to try to show that the components of the current dual vector cannot become too large. Throughout this section, I assume that the initial vector is the null vector; the extension to the case of an arbitrary initial vector is straightforward.

(207) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} that contains a unique entry equal to L :

$$\text{a. } \theta_k^{\text{new}} \doteq \begin{cases} \theta_k^{\text{old}} + 1 & \text{if } k \in W(\mathbf{a}) \\ \theta_k^{\text{old}} - w(\mathbf{a}) & \text{if } k = L(\mathbf{a}) \\ \theta_k^{\text{old}} & \text{otherwise} \end{cases}$$

b. Promote every winner-preferring constraint by 1; demote the loser-preferring constraint by the number $w(\mathbf{a})$ of winner-preferring constraints.

Let me introduce the invariant intuitively, step by step. Consider a generic row \mathbf{a}_i of the input comparative tableau. Let C_ℓ be the unique loser-preferring constraint w.r.t. row \mathbf{a}_i . For concreteness, let me start by assuming that the row \mathbf{a}_i also has a unique winner-preferring constraint C_h . How many times α_i will this row \mathbf{a}_i trigger an update? For concreteness, suppose that \mathbf{a}_i is the first row fed to the algorithm. Since the initial vector is the null vector, the current ranking vector will be OT-compatible with the row \mathbf{a}_i after a single update triggered by it. So, how many further updates will the row \mathbf{a}_i trigger? namely, how can we bound $\alpha_i - 1$? The answer of course depends on how many rows in the input comparative tableau will later on trigger an update that disrupts the ranking

configuration $C_h \gg C_\ell$ necessary for the OT-compatibility with the row a_i . Let me refer to any such row as an ENEMY of row a_i and let me denote by $\text{enemies}(a_i)$ the set of all rows of the input comparative tableau that are enemies of the row a_i . Since only updates triggered by enemies justify further updates by the row a_i , then we expect that the number $\alpha_i - 1$ of updates by the row a_i besides the first one must be bound by the numbers α_j of updates triggered by enemies $a_j \in \text{enemies}(a_i)$ of row a_i multiplied by some proper constant const_j that reflects properties of the specific enemy a_j , as stated in (208).

$$(208) \quad \alpha_i - 1 \leq \sum_{a_j \in \text{enemies}(a_i)} \text{const}_j \cdot \alpha_j$$

What should the constants const_j look like? Recall that the enemies of row a_i are those rows a_j of the input tableau that are able to potentially disrupt the ranking configuration $C_h \gg C_\ell$ required for OT-compatibility with the row a_i . Thus, there are three types of enemies a_j , described in (209): either a_j is an enemy because it pushes down the constraint C_h that a_i pushes up, as in (209a); or because a_j pushes up the constraint C_ℓ that a_i pushes down, as in (209b); or because both situations hold at the same time, as in (209c).

$$(209) \quad \begin{array}{c} \text{a. } a_i \\ a_j \end{array} \begin{bmatrix} \dots & C_h & C_\ell & \dots \\ \dots & W & L & \dots \\ \dots & L & E & \dots \\ \vdots & & & \vdots \end{bmatrix} \quad \begin{array}{c} \text{b. } a_i \\ a_j \end{array} \begin{bmatrix} \dots & C_h & C_\ell & \dots \\ \dots & W & L & \dots \\ \dots & E & W & \dots \\ \vdots & & & \vdots \end{bmatrix} \quad \begin{array}{c} \text{c. } a_i \\ a_j \end{array} \begin{bmatrix} \dots & C_h & C_\ell & \dots \\ \dots & W & L & \dots \\ \dots & L & W & \dots \\ \vdots & & & \vdots \end{bmatrix}$$

We expect the constant const_j to depend on which of the three types of enemies (209) the row a_j is. Let $w(a_j)$ be the number of winner-preferring constraints w.r.t. row a_j . Suppose that a_j is an enemy of type (209a). In this case, row a_j has an L corresponding to constraint C_h ; upon update by a_j according to (207), the ranking value of C_h will decrease by $w(a_j)$; thus, update by a_j disrupts the ranking configuration necessary for a_i by $w(a_j)$. For this reason, it makes sense to let the constant const_j be proportional to $w(a_j)$ in this case. Next, suppose that a_j is an enemy of type (209b). In this case, row a_j has a W corresponding to constraint C_ℓ ; upon update by a_j according to (207), the ranking value of C_ℓ will increase by 1; thus, update by a_j disrupts the ranking configuration necessary for a_i by 1. For this reason, it makes sense to let the constant const_j be proportional to 1 in this case. Finally, suppose that a_j is an enemy of type (209c). In this case, row a_j has a W corresponding to constraint C_ℓ and an L corresponding to constraint C_h ; upon update by a_j according to (207), the ranking value of C_ℓ will increase by 1 and the ranking value of C_h will decrease by $w(a_j)$; thus, update by a_j disrupts the ranking configuration necessary for a_i by $w(a_j) + 1$. For this reason, it makes sense to let the constant const_j be proportional to $w(a_j) + 1$ in this case.

$$(210) \quad \text{const}_j \sim \begin{cases} w(a_j) & \text{if } a_j \text{ is an enemy of type (209a)} \\ 1 & \text{if } a_j \text{ is an enemy of type (209b)} \\ 1 + w(a_j) & \text{if } a_j \text{ is an enemy of type (209c)} \end{cases}$$

So far, I have stuck by the simplifying assumption that the row a_i contains only one entry equal to W, corresponding to the constraint C_h . Now let's remove this assumption, and let $w(a_i)$ be the total number of winner-preferring constraints w.r.t. the row a_i . We expect the discussion so far to hold for each one of these winner-preferring constraints. The set $\text{enemies}(a_i)$ of row a_i will of course depend on the specific winner-preferring constraint considered: a given row a_j can be an enemy, say an enemy of type (209a), w.r.t. to a winner-preferring constraint of row a_i but not with respect to another one. Thus, let me replace the notation $\text{enemies}(a_j)$ with $\text{enemies}(\alpha_j, h)$, that is the set of enemies of row a_i with respect to its winner-preferring constraint $C_h \in W(a_i)$. Furthermore, the constants const_j will of course depend on the specific winner-preferring constraint C_h , since the classification of an enemy into one of the three types (209) depends on the winner-preferring constraint considered. Let me make this dependence explicit, by replacing the notation const_j with const_j^h , as in (211)

$$(211) \quad \text{a. } \alpha_i - 1 \leq \sum_{\mathbf{a}_j \in \text{enemies}(\alpha_i, h)} \text{const}_j^h \cdot \alpha_j$$

$$\text{b. } \text{const}_j^h \sim \begin{cases} w(\mathbf{a}_j) & \text{if } \mathbf{a}_j \text{ is an enemy of type (209a) w.r.t. } C_h \\ 1 & \text{if } \mathbf{a}_j \text{ is an enemy of type (209b) w.r.t. } C_h \\ 1 + w(\mathbf{a}_j) & \text{if } \mathbf{a}_j \text{ is an enemy of type (209c) w.r.t. } C_h \end{cases}$$

Furthermore, once we consider the case of a row \mathbf{a}_i with possibly multiple winner-preferring constraints, then we expect the constants const_j^h to depend on that number, because of the following intuition, already discussed above (123). After update by row \mathbf{a}_i , the separation between the ranking value of the loser-preferring constraint C_ℓ and the ranking value of the winner-preferring constraints will have increased by $w(\mathbf{a}_i) + 1$, since the former is demoted by $w(\mathbf{a}_i)$ and the latter are promoted by 1. If the row \mathbf{a}_i has few winner-preferring constraints, then $w(\mathbf{a}_i)$ is small, hence the gained separation $w(\mathbf{a}_i) + 1$ is small too, hence it will be eroded fast by enemies and row \mathbf{a}_i will therefore require many updates. If instead row \mathbf{a}_i has many winner-preferring constraints, then $w(\mathbf{a}_i)$ is large, hence the gained separation $w(\mathbf{a}_i) + 1$ after an update is large too, hence it will be eroded slowly by enemies and row \mathbf{a}_i will therefore require few updates. These heuristic considerations suggest that the constants const_j^h should also be inversely proportional to the number $w(\mathbf{a}_i) + 1$. Let me make this dependence explicit, by replacing the notation const_j^h with $\text{const}_j^{i,h}$ and by defining the latter constants as in (212).

$$(212) \quad \text{a. } \alpha_i - 1 \leq \sum_{\mathbf{a}_j \in \text{enemies}(\alpha_i, h)} \text{const}_j^{i,h} \cdot \alpha_j$$

$$\text{b. } \text{const}_j^{i,h} = \begin{cases} \frac{w(\mathbf{a}_j)}{w(\mathbf{a}_i) + 1} & \text{if } \mathbf{a}_j \text{ is an enemy of type (209a) w.r.t. } C_h \\ \frac{1}{w(\mathbf{a}_i) + 1} & \text{if } \mathbf{a}_j \text{ is an enemy of type (209b) w.r.t. } C_h \\ \frac{w(\mathbf{a}_j) + 1}{w(\mathbf{a}_i) + 1} & \text{if } \mathbf{a}_j \text{ is an enemy of type (209c) w.r.t. } C_h \end{cases}$$

Claim 27 Consider an input OT-compatible comparative tableau $\mathbf{A} \in \{L, E, W\}^{m \times n}$ whose rows contain exactly one entry equal to L. The dual vector α^t entertained at a generic time t by the on-line algorithm (205) with the promotion-demotion update rule (207) run on this input comparative tableau starting from the null initial vector satisfies the invariant (212), for every row $i = 1, \dots, m$ and every winner-preferring constraint $h \in W(\mathbf{a}_i)$.

Proof. For every row \mathbf{a}_i of the given comparative tableau \mathbf{A} , let $W(\mathbf{a}_i)$ be the set of corresponding winner-preferring constraints, let $L(\mathbf{a}_i)$ be the unique corresponding loser-preferring constraint. Let $\bar{\mathbf{a}}_i$ be the update vector corresponding to the update rule (207) w.r.t. update triggered by the i th row \mathbf{a}_i of the input comparative tableau, as defined in (161) and repeated in (213). Here, I am using the notation $\bar{\mathbf{a}}_i = (\bar{a}_{i,1}, \dots, \bar{a}_{i,h}, \dots, \bar{a}_{i,n})$, so that $\bar{a}_{i,h}$ is the h th entry of the update vector corresponding to the i th comparative row. Let $\bar{\mathbf{A}}$ be the numerical matrix obtained by organising these numerical vectors $\bar{\mathbf{a}}_1, \dots, \bar{\mathbf{a}}_m$ one underneath the other. Let $\bar{\mathbf{A}}_1, \dots, \bar{\mathbf{A}}_h, \dots, \bar{\mathbf{A}}_n$ be the n columns of the numerical matrix $\bar{\mathbf{A}}$.

$$(213) \quad \bar{a}_{i,h} \doteq \begin{cases} 1 & \text{if } h \in W(\mathbf{a}_i) \\ -w(\mathbf{a}_i) & \text{if } h = L(\mathbf{a}_i) \\ 0 & \text{otherwise} \end{cases}$$

For every row $i = 1, \dots, m$ and for every constraint $h \in W(\mathbf{a}_i)$, let $\mathbf{d}^{h,i} = (d_1^{h,i}, \dots, d_m^{h,i}) \in \mathbb{R}^m$ be the vector whose generic j th component $d_j^{h,i}$ is defined as in (214) for every $j = 1, \dots, m$. In words, $\mathbf{d}^{h,i}$ is the difference between the h th column $\bar{\mathbf{A}}_h$ and the column $\bar{\mathbf{A}}_{L(\mathbf{a}_i)}$ of the derived matrix $\bar{\mathbf{A}}$ with all positive components set to zero but the i th one. Thus, by construction we have that $d_j^{h,i}$ is nonpositive for every $j \neq i$; and furthermore that $d_i^{h,i} = w(\mathbf{a}_i) + 1$.

$$(214) \quad d_j^{i,h} \doteq \begin{cases} \bar{a}_{j,h} - \bar{a}_{j,L(\mathbf{a}_i)} & \text{if } a_{j,h} - a_{j,L(\mathbf{a}_i)} \leq 0 \text{ or } j = i \\ 0 & \text{otherwise} \end{cases}$$

With this notation, the invariant (212) can be rewritten compactly as in (215). In the rest of the proof, I show that the inequality (215) holds.

$$(215) \quad \langle \mathbf{d}^{i,h}, \boldsymbol{\alpha}^t \rangle \leq w(\mathbf{a}_i) + 1$$

Fix an arbitrary $i = 1, \dots, m$. If $\alpha_i^t = 1$, then the proof of the invariant (215) for any $h \in W(\mathbf{a}_i)$ consists of the chain of inequalities in (216). Here, I have reasoned as follows: in step (a), I have used the definition (94) of scalar product; in step (b), I have just singled out a special value of the index $j \in \{1, \dots, m\}$, namely the value i ; in step (c), I have upper bounded by dropping the term (*), which is nonpositive because the dual variables are by definition nonnegative and the components of the vector $\mathbf{d}^{i,h}$ in are all nonpositive but for the i th one; in step (d), I have used the hypothesis that $\alpha_i^t = 1$; in step (e), I have used the fact that $d_i^{h,i} = w(\mathbf{a}_i) + 1$.

$$(216) \quad \begin{aligned} \langle \mathbf{d}^{i,h}, \boldsymbol{\alpha}^t \rangle &\stackrel{(a)}{=} \sum_{j=1}^m d_j^{i,h} \alpha_j^t \\ &\stackrel{(b)}{=} d_i^{i,h} \alpha_i^t + \underbrace{\sum_{j=1, j \neq i}^m d_j^{i,h} \alpha_j^t}_{(*)} \\ &\stackrel{(c)}{\leq} d_i^{i,h} \alpha_i^t \\ &\stackrel{(d)}{=} d_i^{i,h} \\ &\stackrel{(e)}{=} w(\mathbf{a}_i) + 1 \end{aligned}$$

If $\alpha_i^t \geq 2$, then the proof of the claim (215) consists of the chain of implications in (217), that completes the proof of the claim.

$$(217) \quad \begin{aligned} \alpha_i^t \geq 2 &\stackrel{(a)}{\implies} \exists s < t \text{ s.t. } \alpha_i^s = \alpha_i^t - 1, \boldsymbol{\theta}^s \text{ not OT-compatible with } \mathbf{a}_i \\ &\stackrel{(b)}{\implies} \exists s < t \text{ s.t. } \alpha_i^s = \alpha_i^t - 1, \theta_h^s \leq \theta_{L(\mathbf{a}_i)}^s \\ &\stackrel{(c)}{\implies} \exists s < t \text{ s.t. } \alpha_i^s = \alpha_i^t - 1, \langle \bar{\mathbf{A}}_h, \boldsymbol{\alpha}^s \rangle \leq \langle \bar{\mathbf{A}}_{L(\mathbf{a}_i)}, \boldsymbol{\alpha}^s \rangle \\ &\stackrel{(d)}{\implies} \exists s < t \text{ s.t. } \alpha_i^s = \alpha_i^t - 1, \langle \bar{\mathbf{A}}_h - \bar{\mathbf{A}}_{L(\mathbf{a}_i)}, \boldsymbol{\alpha}^s \rangle \leq 0 \\ &\stackrel{(e)}{\implies} \exists s < t \text{ s.t. } \alpha_i^s = \alpha_i^t - 1, \langle \mathbf{d}^{i,h}, \boldsymbol{\alpha}^s \rangle \leq 0 \\ &\stackrel{(f)}{\implies} \exists s < t \text{ s.t. } \alpha_i^s = \alpha_i^t - 1, \sum_{j=1}^m d_j^{i,h} \alpha_j^s \leq 0 \\ &\stackrel{(g)}{\implies} \exists s < t \text{ s.t. } \alpha_i^s = \alpha_i^t - 1, d_i^{i,h} \alpha_i^s + \sum_{j=1, j \neq i}^m d_j^{i,h} \alpha_j^s \leq 0 \\ &\stackrel{(h)}{\implies} \exists s < t \text{ s.t. } \alpha_i^s = \alpha_i^t - 1, d_i^{i,h} \alpha_i^s + \sum_{j=1, j \neq i}^m d_j^{i,h} \alpha_j^s \leq 0 \\ &\stackrel{(i)}{\implies} d_i^{i,h} (\alpha_i^t - 1) + \sum_{j=1, j \neq i}^m d_j^{i,h} \alpha_j^t \leq 0 \\ &\implies d_i^{i,h} \alpha_i^t + \sum_{j=1, j \neq i}^m d_j^{i,h} \alpha_j^t \leq d_i^{i,h} \\ &\implies \langle \mathbf{d}^{i,h}, \boldsymbol{\alpha}^t \rangle \leq d_i^{i,h} \\ &\stackrel{(l)}{\implies} \langle \mathbf{d}^{i,h}, \boldsymbol{\alpha}^t \rangle \leq w(\mathbf{a}_i) + 1 \end{aligned}$$

Here, I have reasoned as follows: in step (a), I have noted that if α_i^t has grown to at least 2, then there has got to exist a time s prior to t where update has occurred in response to a failure in accounting

for the corresponding row \mathbf{a}_i ; in step (b), I have noted that the hypothesis that the ranking vector θ^s at time s is not OT-compatible with the row \mathbf{a}_i entails in particular that the winner-preferring constraint C_h is not ranked above the unique loser-preferring constraint $C_{L(\mathbf{a}_i)}$ or equivalently that $\theta_h^s \leq \theta_{L(\mathbf{a}_i)}^s$; in step (c), I have expressed the primal variables θ_h^s and $\theta_{L(\mathbf{a}_i)}^s$ in terms of the dual variables, using the corresponding columns $\overline{\mathbf{A}}_h$ and $\overline{\mathbf{A}}_{L(\mathbf{a}_i)}$ of the derived numerical matrix, as in (320); in step (d), I have used linearity of the scalar product; in step (e), I have used the fact that the vector $\mathbf{d}^{i,h}$ coincides with the vector $\overline{\mathbf{A}}_h - \overline{\mathbf{A}}_{L(\mathbf{a}_i)}$ with the positive components set to zero, and thus $\langle \mathbf{d}^{i,h}, \boldsymbol{\alpha}^s \rangle \leq \langle \overline{\mathbf{A}}_h - \overline{\mathbf{A}}_{L(\mathbf{a}_i)}, \boldsymbol{\alpha}^s \rangle$; in step (f), I have used the definition (94) of scalar product; in step (g), I have singled out one specific value of the index $j \in \{1, \dots, n\}$, namely $j = i$; in step (h), I have used the fact that $s < t$ implies that $\alpha_j^s \leq \alpha_j^t$ (because the dual variables can only increase over time) and thus that $d_j^{i,h} \alpha_j^s \geq d_j^{i,h} \alpha_j^t$ (because the components of the vector $\mathbf{d}^{i,h}$ are nonpositive by construction); in step (i), I have just rewritten α_i^s as $\alpha_i^t - 1$; in the last step (l), I have used the fact that $d_i^{h,i} = w(\mathbf{a}_i) + 1$. ■

As noted at the beginning of the chapter, the dual invariant provided by claim 27 applies in particular to the final dual vector. Furthermore, the final dual vector determines the final primal vector, as noted in subsection 4.2.6. In conclusion, the invariant provided by claim 27 offers an indirect characterization of the final ranking vector returned by the algorithm, thus tackling question (206b). Yet, the invariant provided by claim 27 is rather weak, and thus the corresponding characterization of the final vector is rather loose. Here is a way to illustrate this point. Let a DIAGONAL TABLEAU of order n be a tableau with the form (56), namely a tableau with n columns and $n - 1$ rows, whose entries are all equal to E but for the i th entry in the i th row that is equal to w and the entry immediately to its right that is equal to L . Let a PATER'S TABLEAU of order n be a tableau with the form (??), namely a diagonal tableau with an extra entry equal to w at the immediate right of the entry equal to L in the first $n - 2$ rows. Finally, consider the variant of Pater's tableau obtained by adding two entries equal to w at the immediate right of the entry equal to L in the first $n - 3$ rows, only one in the penultimate row and none in the last row. I illustrate in (218) the three types of comparative tableaux for the case $n = 5$.

$$(218) \quad \begin{array}{l} \text{a.} \\ \text{b.} \\ \text{c.} \end{array} \begin{array}{c} \begin{array}{ccccc} C_1 & C_2 & C_3 & C_4 & C_5 \\ \left[\begin{array}{ccccc} W & L & & & \\ & W & L & & \\ & & W & L & \\ & & & W & L \\ & & & & W & L \end{array} \right] \\ \\ \begin{array}{ccccc} C_1 & C_2 & C_3 & C_4 & C_5 \\ \left[\begin{array}{ccccc} W & L & W & & \\ & W & L & W & \\ & & W & L & W \\ & & & W & L \end{array} \right] \\ \\ \begin{array}{ccccc} C_1 & C_2 & C_3 & C_4 & C_5 \\ \left[\begin{array}{ccccc} W & L & W & W & \\ & W & L & W & W \\ & & W & L & W \\ & & & W & L \end{array} \right] \end{array} \end{array}$$

Multiple simulations show that the final vector entertained by the on-line algorithm (205) with the update rule (207) run on any of these three families of comparative tableaux with an odd number n of columns always has the shape in (219), no matter the order in which the rows are fed to the algorithm.

$$(219) \quad \begin{aligned} \theta_1^{\text{fin}} &= \frac{n-1}{2} \\ &\vdots \\ \theta_{\frac{n-1}{2}}^{\text{fin}} &= 0 \\ &\vdots \\ \theta_n^{\text{fin}} &= -\frac{n-1}{2} \end{aligned}$$

In other words, the final vector θ^{fin} entertained by the on-line algorithm (205) run with the promotion-demotion update rule (207) on any of the three families of tableaux exemplified in (218) with odd n can be uniquely characterized as in (220). Yet, the characterization of the final vector obtained through the invariant of claim 27 does not in any way provide such a sharp characterization.¹

$$(220) \quad \theta^{\text{fin}} = \min \left\{ \|\theta\|_1 = \sum_{h=1}^n |\theta_h| \mid \theta \in \mathbb{Z}^n, \theta \text{ OT-compatible with } \mathbf{A} \right\}$$

The reason why the invariant provided in claim 27 is loose is rather obvious. Given a row \mathbf{a}_i of the input tableau, this invariant only considers rows \mathbf{a}_j that are enemies of that row, namely that potentially disrupt the ranking configuration necessary for OT-compatibility with the row \mathbf{a}_i . This invariant does not take into account rows \mathbf{a}_j that help establishing the ranking configuration necessary for OT-compatibility with the row \mathbf{a}_i . Let me call these rows HELPERS. Analogously to (209), an helper of row \mathbf{a}_i is a row \mathbf{a}_j of one of the three types illustrated in (221): either both \mathbf{a}_i and \mathbf{a}_j push up the constraint C_h , as in (221a); or both \mathbf{a}_i and \mathbf{a}_j push down the constraint C_ℓ , as in (221b); or both situations hold at the same time, as in (221c).

$$(221) \quad \begin{array}{c} \mathbf{a}_i \\ \mathbf{a}_j \end{array} \begin{bmatrix} \dots & C_h & C_\ell & \dots \\ \vdots & & & \\ \dots & W & L & \dots \\ \dots & W & E & \dots \\ \vdots & & & \end{bmatrix}, \quad \begin{array}{c} \mathbf{a}_i \\ \mathbf{a}_j \end{array} \begin{bmatrix} \dots & C_h & C_\ell & \dots \\ \vdots & & & \\ \dots & W & L & \dots \\ \dots & E & L & \dots \\ \vdots & & & \end{bmatrix}, \quad \begin{array}{c} \mathbf{a}_i \\ \mathbf{a}_j \end{array} \begin{bmatrix} \dots & C_h & C_\ell & \dots \\ \vdots & & & \\ \dots & W & L & \dots \\ \dots & W & L & \dots \\ \vdots & & & \end{bmatrix}$$

Let $\text{helpers}(\alpha_i, h)$ be the set of helpers of row \mathbf{a}_i relatively to a winner-preferring constraint $C_h \in W(\mathbf{a}_i)$. Of course, we expect the number α_i of updates triggered by row \mathbf{a}_i to depend also on the number of updates α_j triggered by helper rows \mathbf{a}_j , in the sense that if the helpers $\mathbf{a}_j \in \text{helpers}(\alpha_i, h)$ trigger many updates, then the row \mathbf{a}_i will trigger few updates. In other words, we expect that the invariant (212) should be strengthened into something like (222), for some properly defined constants $\text{const}_j^{i,h}$ for $\mathbf{a}_j \in \text{helpers}(\alpha_i, h)$.

$$(222) \quad \alpha_i - 1 \leq \sum_{\mathbf{a}_j \in \text{enemies}(\alpha_i, h)} \text{const}_j^{i,h} \cdot \alpha_j - \sum_{\mathbf{a}_j \in \text{helpers}(\alpha_i, h)} \text{const}_j^{i,h} \cdot \alpha_j$$

But so far I have not been able to prove anything like (222).

6.2 Preliminaries on OT phonotactics

Given a language \mathcal{L} in the typology corresponding to some universal specifications $(\mathcal{X}, \mathcal{Y}, \text{Gen}, C)$, let me say that a ranking \gg OT-CORRESPONDS to the phonotactics of the target language \mathcal{L} iff it satisfies condition (223), namely the language generated by the corresponding OT-grammar OT_{\gg} coincides with the target language \mathcal{L} . In this section, I am interested in the following problem: given a target language \mathcal{L} , how can I characterize the rankings that OT-correspond to it? I will discuss this problem under various restrictive assumptions on the universal specifications $(\mathcal{X}, \mathcal{Y}, \text{Gen}, C)$. The investigation of this problem will turn out very useful in the discussion of the various case studied considered in the next section.

¹It is trivial to prove that (219) does indeed hold for diagonal tableaux (218a). But to understand why it holds in the case of the other two families of comparative tableaux (218b) and (218c) is not trivial.

$$(223) \quad \mathcal{R}(\text{OT}_{\gg}) = \mathcal{L}$$

Let me start by introducing a first set of assumptions on the universal specifications $(\mathcal{X}, \mathcal{Y}, \text{Gen}, \mathcal{C})$. I assume that the set of underlying forms \mathcal{X} and the set of surface forms \mathcal{Y} coincide, as stated in (224a). This assumption makes sense for many realistic cases, although admittedly not for cases such as syllabification, stress assignment, and so on, where the set of output forms is richer than the set of input forms. Furthermore, I assume that a given surface form $y \in \mathcal{Y}$ belongs to the language $\mathcal{R}(\text{OT}_{\gg})$ corresponding to a ranking \gg iff the corresponding grammar OT_{\gg} maps that form y , construed as an underlying form by virtue of (224a), faithfully into itself, as stated in (224b). Tesar (2008) calls any grammar that satisfies this condition an OUTPUT-DRIVEN MAP. Furthermore, he proves that OT-grammars are output-driven maps under mild assumptions on the constraint set \mathcal{C} .

$$(224) \quad \begin{array}{l} \text{a. } \mathcal{X} = \mathcal{Y} \\ \text{b. } y \in \mathcal{R}(\text{OT}_{\gg}) \iff \text{OT}_{\gg}(y) = y \end{array}$$

Given the two assumptions in (224), I can pair up a language \mathcal{L} with the corresponding PHONOTACTIC COMPARATIVE TABLEAU defined as the comparative tableau obtained by listing one underneath the other the comparative rows corresponding to all underlying/winner/loser form triplets (y, y, z) for all underlying/winner forms $y \in \mathcal{L}$ in the target language and all corresponding loser candidates $z \in \text{Gen}(y) \setminus \{y\}$. I will denote by $\mathbf{A}_{\mathcal{L}}$ the phonotactic comparative tableau corresponding to a language \mathcal{L} . A phonotactic comparative tableau is made up of two blocks: the block corresponding to the faithfulness constraints, that does not contain a single entry equal to L; and the block corresponding to the markedness constraints, that can contain entries equal to L. I will call any row of a phonotactic comparative tableau a PHONOTACTIC COMPARATIVE ROW.

$$(225) \quad \mathbf{A}_{\mathcal{L}} = \left[\begin{array}{c|c} \text{faithfulness constraints} & \text{markedness constraints} \\ \hline & \\ \hline \text{E, W} & \text{L, E, W} \\ \hline & \\ \hline \end{array} \right]$$

Of course, a ranking \gg is OT-compatible with the phonotactic comparative tableau $\mathbf{A}_{\mathcal{L}}$ corresponding to a language \mathcal{L} iff $\mathcal{L} \subseteq \mathcal{R}(\text{OT}_{\gg})$, namely the target language \mathcal{L} is a subset of the language corresponding to the ranking \gg . In order to establish (223), we thus need to worry about the reverse inclusion $\mathcal{L} \supseteq \mathcal{R}(\text{OT}_{\gg})$. To this end, let me introduce some more assumptions on the universal specifications $(\mathcal{X}, \mathcal{Y}, \text{Gen}, \mathcal{C})$. I assume that the generating function Gen , construed as a relation on $\mathcal{X} = \mathcal{Y}$, is symmetric, as stated in (224c). This assumption makes sense if the set $\text{Gen}(x)$ of candidates for x is construed as the set of those forms that can be obtained from x by performing reversible operations (epenthesis, deletion, modification of the value of a feature, etcetera). Furthermore, I assume that all faithfulness constraints are binary, namely either $F(x, y) = 0$ (if the pair (x, y) does not differ along the dimension relevant to F) or else $F(x, y) = 1$ (if the pair (x, y) does differ along the dimension relevant to F). Thus, a faithfulness constraint F can be construed as a relation over $\mathcal{X} \times \mathcal{Y}$. Since $\mathcal{X} = \mathcal{Y}$ by assumption (224a), then I can assume that this relation is symmetric for all faithfulness constraints F , as stated in (224d). This assumption makes sense for (binary) faithfulness constraints of the IDENTITY type, but does not hold for DEP and MAX. All four assumptions (224) will hold for the test cases studied in the next section.

$$(224) \quad \begin{array}{l} \text{c. } y \in \text{Gen}(x) \text{ iff } x \in \text{Gen}(y) \\ \text{d. } F(x, y) = F(y, x) \end{array}$$

Each row of a phonotactic comparative tableau corresponds to a underlying/winner/loser form triplet (y, y, z) , for some $y \in \mathcal{L}$ and some $z \in \text{Gen}(y)$. Thus, there are two different, interesting ways to organize the rows of a phonotactic comparative tableau: by their underlying/winner form y ; or by their loser form z . In this section, I use the latter ordering. Thus, let me denote by $\mathbf{A}_{(\cdot, \cdot, z)}$ the block of all those rows of the phonotactic comparative tableau that correspond to the triplets (y, y, z) for some $y \in \mathcal{L}$ such that $z \in \text{Gen}(y)$, as in (226).

$$(226) \quad \mathbf{A}_{(\cdot, \cdot, z)} = \left\{ (y, y, z) \mid y \in \mathcal{L}, z \in \text{Gen}(y) \setminus \{y\} \right\}$$

A phonotactic comparative tableau can thus be reorganized as in (227), by ordering its rows into the different blocks $\mathbf{A}_{(\cdot, \cdot, z)}$. Note that this blocks can be classified into two types: the blocks $\mathbf{A}_{(\cdot, \cdot, z)}$ corresponding to a loser candidate z that belongs to the target language \mathcal{L} and the blocks $\mathbf{A}_{(\cdot, \cdot, z)}$ corresponding to a loser candidate z that does not belong to the target language \mathcal{L} .

	faithfulness constraints	markedness constraints
(227) $\mathbf{A}_{(\cdot, \cdot, z_1)}$	E, W	L, E, W
$\mathbf{A}_{(\cdot, \cdot, z_2)}$	E, W	L, E, W
\vdots	E, W	L, E, W

The universal specifications in (1) in subsection 4.1.1 satisfy all four assumptions (224). Consider the corresponding language (7), repeated in (228a). The only ranking that satisfies condition (223) and thus OT-corresponds to this language is the ranking in (3), repeated in (228b). The phonotactic comparative tableau corresponding to the language \mathcal{L} in (228a) is given in (228c). We can think of the three rows of this comparative tableau as three different blocks $\mathbf{A}_{(\cdot, \cdot, z)}$, namely the three blocks $\mathbf{A}_{(\cdot, \cdot, \text{ta})}$, $\mathbf{A}_{(\cdot, \cdot, \text{da})}$ and $\mathbf{A}_{(\cdot, \cdot, \text{rat})}$, with the irrelevant peculiarity that each block contains a unique row. The ranking (228b) can be characterized in terms of the phonotactics tableau (228c) by means of the following two conditions. *First*, it is OT-compatible with the phonotactic comparative tableau (228c), namely it satisfies the condition that either F_{pos} or F_{gen} are ranked above M . *Second*, it has the property that the decisive constraint corresponding to the third row is a markedness constraint. Note that the third row is the only one among the three blocks $\mathbf{A}_{(\cdot, \cdot, z)}$ that make up the phonotactic comparative tableau (228c) that corresponds to a loser form $z = [\text{rad}]$ that does not belong to the target language \mathcal{L} in (228a). The following claim says that this characterization of OT-corresponding rankings that we have devised in the special case of (228) actually extends to the general case.

- (228) a. $\mathcal{L} = \{[\text{ta}], [\text{da}], [\text{rat}]\}$
 b. $F_{\text{pos}} \gg M \gg F_{\text{gen}}$

c.	$\begin{array}{l} (\text{da}, \text{da}, \text{ta}) \\ (\text{ta}, \text{ta}, \text{da}) \\ (\text{rat}, \text{rat}, \text{rad}) \end{array}$	$\left[\begin{array}{ccc} F_{\text{pos}} & F_{\text{gen}} & M \\ \hline \text{W} & \text{W} & \text{L} \\ \hline \text{W} & \text{W} & \text{W} \\ \hline \text{E} & \text{W} & \text{W} \end{array} \right]$	← only block $\mathbf{A}_{(\cdot, \cdot, z)}$ corresponding to $z \notin \mathcal{L}$
----	---	--	---

Claim 28 Consider universal specifications $(\mathcal{X}, \mathcal{Y}, \text{Gen}, \mathcal{C})$ that satisfy all four assumptions (224). Consider an entire language \mathcal{L} in the corresponding typology. A ranking \gg satisfies condition (223) and thus OT-corresponds to the target language \mathcal{L} iff \gg is OT-compatible with the phonotactics tableau $\mathbf{A}_{\mathcal{L}}$ and furthermore for every $y \in \mathcal{Y} \setminus \mathcal{L}$ there exists at least one row \mathbf{a} in the block $\mathbf{A}_{(\cdot, \cdot, y)}$ such that $\text{dec}_{\gg}(\mathbf{a}) \in \mathcal{M}$.

Proof. As noted above, assumption (224b) entails that $\mathcal{R}(\text{OT}_{\gg}) \supseteq \mathcal{L}$ iff \gg is OT-compatible with the phonotactic comparative tableau $\mathbf{A}_{\mathcal{L}}$. Thus, consider one such ranking \gg . The proof of the claim then consists of the logical equivalences in (229). Here, I have reasoned as follows: in step (a), I have used the hypothesis that \gg is OT-compatible with the phonotactic comparative tableau corresponding to \mathcal{L} and thus $\mathcal{L} \subseteq \mathcal{R}(\text{OT}_{\gg})$; in step (b), I have used the fact that $y \in \text{dom}(\text{OT}_{\gg})$, by RICHNESS OF THE BASE; steps (c) and (d) are less trivial, and are explained in detail below; step (e) is just a restatement.

with. A markedness constraint M has a W (or a E or an L) in the row corresponding to (x, x, y) iff it has an L (or a E or a W, respectively) in the row corresponding to (y, y, x) . Furthermore, assumption (224d) ensures that a faithfulness constraint F has a W (a E) in the row corresponding to (x, x, y) iff it has a W (a E, respectively) in the row corresponding to (y, y, x) . The relationship between the rows corresponding to the two triplets (x, x, y) and (y, y, x) is thus as depicted in (231).

$$(231) \quad \begin{array}{c} \begin{array}{cccccc} & F & & F' & & M & & M' & & M'' \\ (x,x,y) & [\dots & W & \dots & E & \dots & L & \dots & E & \dots & W & \dots] \end{array} \\ \Downarrow \\ (y,y,x) & [\dots & W & \dots & E & \dots & W & \dots & E & \dots & L & \dots] \end{array}$$

Suppose that the \gg -decisive winner-preferring constraint in the row corresponding to (x, x, y) is a markedness constraint. This means that there exists a markedness constraint M that has a W in the row corresponding to (x, x, y) and furthermore every constraint \gg -ranked above M has a E in that row (x, x, y) . This means in turn that that markedness constraint M has an L in the row corresponding to (y, y, x) and furthermore every constraint \gg -ranked above M has a E in that row (x, x, y) . Since the latter row (x, x, y) belongs to the phonotactic comparative tableau corresponding to the target language, then the ranking \gg is not OT-compatible with that tableau, contradicting the hypothesis that it is. ■

Given a ranking \gg over the constraint set \mathcal{C} , let me define the sub-relation \gg_{im} of \gg as in (232). If $C \gg_{\text{im}} C'$, then I will say that C is ranked IMMEDIATELY above C' . For example, given the ranking $C_1 \gg C_2 \gg C_3$, we have that $C_1 \gg_{\text{im}} C_2$ but that $C_1 \not\gg_{\text{im}} C_3$. Obviously, \gg is the transitive closure of \gg_{im} .

$$(232) \quad \text{We have } C \gg_{\text{im}} C' \text{ iff } C \gg C' \text{ and there exists no } C'' \text{ such that } C \gg C'' \gg C'.$$

It is useful to concentrate on immediate rankings, because it is easier to study what happens when we revert them. Given a ranking \gg that contains the immediate ranking $C \gg_{\text{im}} C'$, let $\gg_{[C,C']}$ be the ranking identical to \gg but for the fact that the two constraints C and C' have been swamped. We say that the immediate ranking $C \gg_{\text{im}} C'$ of \gg is CRUCIAL w.r.t. \mathcal{L} iff \gg OT-corresponds to the target language \mathcal{L} but $\gg_{[C,C']}$ does not, namely the two conditions (233) hold.

$$(233) \quad \begin{array}{l} \text{a. } \mathcal{R}(\text{OT}_{\gg}) = \mathcal{L} \\ \text{b. } \mathcal{R}(\text{OT}_{\gg_{[C,C']}}) \neq \mathcal{L} \end{array}$$

The preceding claim 28 entails the following characterization of crucial immediate rankings. This simple claim will turn out very useful in the discussion of the test case in subsection 6.3.3.

Claim 30 Consider a language \mathcal{L} and a ranking \gg that OT-corresponds to it, namely such that $\mathcal{R}(\text{OT}_{\gg}) = \mathcal{L}$. If an immediate ranking $M \gg_{\text{im}} F$ of a markedness constraint M above a faithfulness constraint F is crucial in \gg w.r.t. \mathcal{L} , then there exists a block $\mathbf{A}_{(\cdot, \cdot, y)}$ in the phonotactic comparative tableau corresponding to the target language \mathcal{L} such that the properties in (234) hold.

$$(234) \quad \begin{array}{l} \text{a. } y \notin \mathcal{L} \\ \text{b. there is a row } \mathbf{a} \text{ in the block } \mathbf{A}_{(\cdot, \cdot, y)} \text{ such that } \text{dec}_{\gg}(\mathbf{a}) = M \text{ and } F \text{ has a w in } \mathbf{a}; \\ \text{c. for any other row } \mathbf{a}' \text{ in the block } \mathbf{A}_{(\cdot, \cdot, y)}, \text{ we have } \text{dec}_{\gg}(\mathbf{a}') \in \mathcal{F}. \end{array}$$

If an immediate ranking $F \gg_{\text{im}} M$ of a faithfulness constraint F above a markedness constraint M is crucial in \gg w.r.t. \mathcal{L} , then there exists a row \mathbf{a} in the phonotactic comparative tableau corresponding to the target language \mathcal{L} such that the property (235) holds.

$$(235) \quad \text{dec}_{\gg}(\mathbf{a}) = F \text{ and } M \text{ has an L in } \mathbf{a}$$

Finally, the immediate ranking $F \gg_{\text{im}} F'$ of a faithfulness constraint F above another faithfulness constraint F' can never be crucial in \gg w.r.t. \mathcal{L} .

6.3 Test cases

In section 6.1, I have tackled question (206b) by looking for a characterization of the final vector entertained by the on-line algorithm (205) with a promotion-demotion update rule that holds in the general case. In this section, I switch gears and turn to the study of the final vector in a few test cases. In section 4.3, I have motivated the need for promotion-demotion update rules for the on-line algorithm (205) by means of cases where the input tableau is a phonotactics tableau. It thus makes sense now to consider test cases where the input tableau is indeed a phonotactics comparative tableau. All test cases considered in this section correspond to universal specifications that satisfy the four assumptions (224). Thus, the target ranking corresponding to a comparative tableau can be characterized as in (236), by virtue of claim 28.

- (236) a. \gg is OT-compatible with the input phonotactics comparative tableau;
 b. every block $\mathbf{A}_{(\cdot, \cdot, z)}$ that corresponds to a loser form z that does not belong to the target language contains at least a row whose \gg -decisive constraint is a markedness constraint.

Following the discussion of section 4.3, I will assume a BIASED initial ranking vector θ^{init} whereby the markedness constraints start out with a ranking value larger than the ranking value of the faithfulness constraints. More in detail, if the constraint set only contains general faithfulness constraints, then I will assume the initial ranking vector described in (237a): the faithfulness constraints all start out with an initial null ranking value; the markedness constraints all start out with an initial ranking value equal to a given positive constant θ^{init} . If the constraint set contains both general and positional faithfulness constraints, then I will assume the initial ranking vector described in (237b): the general faithfulness constraints start out with a null initial ranking value; the positional faithfulness constraints start out with an initial ranking value equal to a positive constant θ^{init} ; the markedness constraints start out with an initial ranking value equal to twice that positive constant. The choice of the actual ratio between the different initial ranking values is completely arbitrary. The advantage of the choice in (237) is that the initial vector is determined by a unique constant θ^{init} . Note that this choice of the initial ranking vector ensures that the current ranking vector has nonnegative components at every time.

- (237) For some positive constant $\theta^{\text{init}} > 0$
 a. $\theta_F^{\text{init}} = 0, \theta_M^{\text{init}} = \theta^{\text{init}}$
 b. $\theta_{F_{\text{gen}}}^{\text{init}} = 0, \theta_{F_{\text{pos}}}^{\text{init}} = \theta^{\text{init}}, \theta_M^{\text{init}} = 2\theta^{\text{init}}$

As noted in section 4.3, we cannot adopt a demotion-only update rule in step 3 of algorithm (205) in the case of an input phonotactics comparative tableau. In fact, since the faithfulness constraints are never loser-preferring, their ranking values would never be modified by a demotion-only update rule. Thus, I need to use a promotion-demotion update rule in step 3. Among the arsenal of convergent promotion-demotion update rules devised in chapter 5, I will adopt the one in (238) in the following test cases. The rationale for this choice is that the promotion amount is always 1, and thus yields a simple dynamics that is easier to study with paper and pencil. Convergence of this update rule is ensured by both lines of analysis presented in chapter 5. It is ensured by the line of analysis of section 5.2, as noted in subsection 5.3.2. And it is also ensured by the line of analysis of section 5.4, since the update rule satisfies all hypotheses of claim 22 — in particular it satisfies hypothesis (175b), since the ranking values of all constraints stay nonnegative throughout learning.

- (238) If the current ranking vector θ^{old} is not OT-compatible with the comparative row \mathbf{a} :
- a. $\theta_k^{\text{new}} \doteq \begin{cases} \theta_k^{\text{old}} + 1 & \text{if } k \in W(\mathbf{a}) \\ \theta_k^{\text{old}} - w(\mathbf{a}) & \text{if } k = L(\mathbf{a}) \\ \theta_k^{\text{old}} & \text{otherwise} \end{cases}$
- b. Promote every winner-preferring constraint by 1; demote the unique loser-preferring constraint by the total number $w(\mathbf{a})$ of winner-preferring constraints.

Finally, I will assume that the rows of the input phonotactics comparative tableau are sampled uniformly. The simulations of the on-line algorithm (205) with the promotion-demotion update rule (238) and the implementation details just described have been performed using the Matlab file `AdditiveGLA`, available on the author's website.

6.3.1 First test case

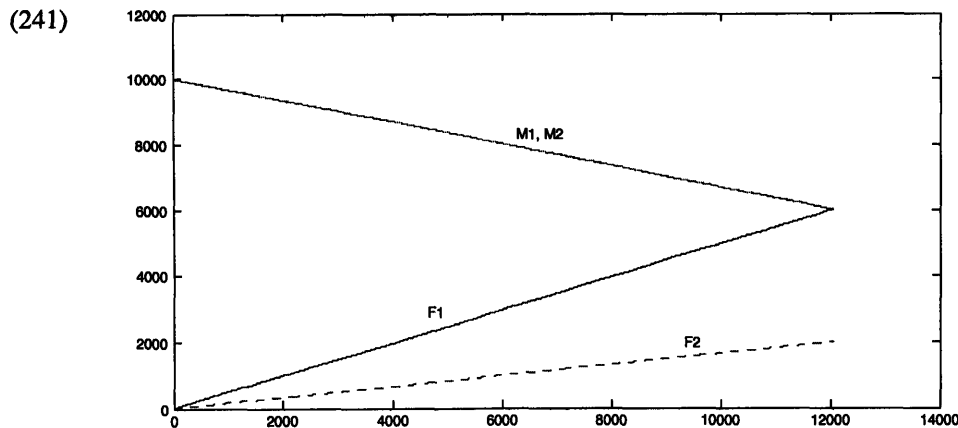
Let me start with the artificial, very simple case of the input comparative tableau in (239). Assume that each row corresponds to a different block $A_{(\cdot, \cdot, z)}$ and that the last two blocks correspond to loser forms that do not belong to the target language. By (236), the target ranking is therefore $F_1 \gg M_1 \gg M_2 \gg F_2$.

$$(239) \quad \begin{array}{c} F_1 \quad F_2 \quad M_1 \quad M_2 \\ \hline W \qquad \qquad L \\ \hline \qquad W \quad W \quad L \\ \hline \qquad W \quad W \\ \hline \qquad W \qquad \qquad W \end{array}$$

The last two rows of the comparative tableau (239) never trigger any update. In order to study the behavior of the algorithm on this input tableau, I can thus get rid of these last two rows, and simplify the tableau as in (240). This case is also considered in Prince and Tesar (2004).

$$(240) \quad A = \begin{array}{c} F_1 \quad F_2 \quad M_1 \quad M_2 \\ \hline W \qquad \qquad L \\ \hline \qquad W \quad W \quad L \end{array}$$

To get a sense of what happens in the case of the input comparative tableau (240), consider the case where the two rows of the tableau are sampled uniformly. The dynamics over time of the ranking values of the four constraints entertained by the on-line algorithm (205) with the promotion-demotion update rule (238) starting from the biased initial vector (237a) with $\theta^{\text{init}} = 10,000$ is reported in (241). The diagram shows two crucial facts: *first*, the two markedness constraints M_1 and M_2 sink with approximately the same speed, as shown by their overlapping trajectories; *second*, the two faithfulness constraints F_1 and F_2 raise with different speeds, namely F_1 raises faster than F_2 . Thus, at the time M_1 and M_2 both cross F_1 , the latter constraint is well above F_2 . From this moment on, the first row of the comparative tableau (240) is accounted for, and will therefore not trigger any further update. Furthermore, since the two markedness constraints are so close, M_1 will get above M_2 much faster than it would take for M_2 to drop below F_2 . From the moment M_1 gets above M_2 on, also the second row of the comparative tableau (240) is accounted for, and the algorithm does not perform any further update. Since nothing requires the two markedness constraints to further sink and cross F_2 , they end up in between the two faithfulness constraints, as desired.



Why is it that the two markedness constraints M_1 and M_2 stay so close? and why is it that F_1 raises faster than F_2 ? To start getting a grasp on these two questions, let's approximate random sampling

of the two rows of the comparative tableau (240) with a deterministic sampling whereby the first row is sampled first, then the second row, than again the first row, and so on. The beginning of the corresponding primal path described by the algorithm is given in (242). Here I only provide the ranking values of the two markedness constraints M_1 and M_2 , and I leave the two ranking values blank when the current ranking vector is OT-compatible with the current comparative row and thus no update is performed. We see in (242) that row 2 triggers an update one third of the times it is fed to the algorithm, while row 1 always does trigger an update. Since F_1 is only pushed up by the first row and F_2 is only pushed up by the second row, then we expect F_1 to raise three times as fast as F_2 . This agrees well with the simulation in (241). Furthermore, we see in (242) that the ranking values of the two markedness constraints differ over time by at most 3, namely they are basically identical. Again, this agrees well with the simulation in (241).

$$(242) \quad \begin{array}{c} \left[\begin{array}{c} 100 \\ 100 \end{array} \right] \xrightarrow{\text{row 1}} \left[\begin{array}{c} 99 \\ 100 \end{array} \right] \xrightarrow{\text{row 2}} \left[\begin{array}{c} 100 \\ 98 \end{array} \right] \xrightarrow{\text{row 1}} \left[\begin{array}{c} 99 \\ 98 \end{array} \right] \xrightarrow{\text{row 2}} \left[\begin{array}{c} \\ \\ \end{array} \right] \xrightarrow{\text{row 1}} \left[\begin{array}{c} 98 \\ 98 \end{array} \right] \xrightarrow{\text{row 2}} \left[\begin{array}{c} 99 \\ 96 \end{array} \right] \\ \xrightarrow{\text{row 1}} \left[\begin{array}{c} 98 \\ 96 \end{array} \right] \xrightarrow{\text{row 2}} \left[\begin{array}{c} \\ \\ \end{array} \right] \xrightarrow{\text{row 1}} \left[\begin{array}{c} 97 \\ 96 \end{array} \right] \xrightarrow{\text{row 2}} \left[\begin{array}{c} \\ \\ \end{array} \right] \xrightarrow{\text{row 1}} \left[\begin{array}{c} 96 \\ 96 \end{array} \right] \xrightarrow{\text{row 2}} \left[\begin{array}{c} 97 \\ 94 \end{array} \right] \\ \xrightarrow{\text{row 1}} \left[\begin{array}{c} 96 \\ 94 \end{array} \right] \xrightarrow{\text{row 2}} \left[\begin{array}{c} \\ \\ \end{array} \right] \xrightarrow{\text{row 1}} \left[\begin{array}{c} 95 \\ 94 \end{array} \right] \xrightarrow{\text{row 2}} \left[\begin{array}{c} \\ \\ \end{array} \right] \xrightarrow{\text{row 1}} \left[\begin{array}{c} 94 \\ 94 \end{array} \right] \xrightarrow{\text{row 2}} \left[\begin{array}{c} 95 \\ 92 \end{array} \right] \end{array}$$

So far, I have crucially relied on assumptions on how the two rows of the comparative tableau (240) are sampled and fed to the algorithm: either uniformly or in a fixed order one after the other. Quite surprisingly, it turns out that the algorithm ends up in the desired ranking no matter how the rows are fed to the algorithm, as stated by the following claim 31. The proof hinges on the invariant presented in section 6.1. An intuitive explanation of the proof concludes this section.

Claim 31 The on-line algorithm (205) with the promotion-demotion update rule (238), starting from the initial vector (237a), run on on the input comparative tableau (240) converges to a final ranking vector θ that represents the desired ranking $F_1 \gg M_1 \gg M_2 \gg F_2$, no matter how the two rows are sampled, provided that the initial ranking value θ^{init} of the markedness constraints is large enough.

Proof. We know that the on-line algorithm (205) with the promotion-demotion update rule (238) converges to a ranking vector OT-compatible with the input tableau (240). Thus, OT-compatibility with the first row of the input tableau entails that the final ranking vector satisfies the strict inequality $\theta_{F_1}^{\text{fin}} > \theta_{M_1}^{\text{fin}}$, that corresponds to the immediate ranking (243a). Suppose furthermore that the final ranking vector satisfies the strict inequality $\theta_{M_2}^{\text{fin}} > \theta_{F_2}^{\text{fin}}$, that corresponds to the immediate ranking (243c). In this case, OT-compatibility with the second row of the input tableau (240) entails that the final vector must satisfy the strict inequality $\theta_{M_1}^{\text{fin}} > \theta_{M_2}^{\text{fin}}$ that corresponds to the immediate ranking (243b). In conclusion, to prove the claim it is sufficient to show that the final ranking vector satisfies the inequality $\theta_{M_2}^{\text{fin}} > \theta_{F_2}^{\text{fin}}$.

$$(243) \quad \begin{array}{ccccccc} F_1 & \gg & M_1 & \gg & M_2 & \gg & F_2 \\ & & | & & | & & | \\ & & \text{(a)} & & \text{(b)} & & \text{(c)} \end{array}$$

Let $\alpha^t = (\alpha_1^t, \alpha_2^t)$ be the dual vector entertained by the algorithm at a generic time t , namely α_1^t (respectively, α_2^t) is the number of times that the algorithm has updated its current ranking vector in response to a failure in accounting for the first (respectively, the second) row of the comparative tableau (240). The ranking values of the two constraints F_2 and M_2 at a generic time t can then be expressed in terms of the dual variables at that same time t as in (244).

$$(244) \quad \begin{array}{l} \theta_{F_2}^t = \alpha_2^t \\ \theta_{M_2}^t = -2\alpha_2^t + \theta^{\text{init}} \end{array}$$

Thus, the desired primal strict inequality $\theta_{M_2}^t > \theta_{F_2}^t$ holds at a given time t iff the dual strict inequality (245) holds at that same time. In the rest of the proof, I will show that the inequality (245) does indeed hold at any time t .

$$(245) \quad 3\alpha_2^t < \theta^{\text{init}}$$

To this end, let me start by noting that the dual vector $\alpha^t = (\alpha_1^t, \alpha_2^t)$ entertained by the algorithm at a generic time t satisfies the two inequalities (246); I will then derive the strict inequality (245) from these two inequalities (246).

$$(246) \quad \begin{aligned} \text{a. } & \alpha_1^t \leq 1 + \frac{1}{2}(\alpha_2^t + \theta^{\text{init}}). \\ \text{b. } & \alpha_2^t \leq \min \left\{ 1 + \frac{1}{3}\theta^{\text{init}}, 1 + \frac{1}{3}\alpha_1^t \right\}. \end{aligned}$$

The two inequalities (246) are an instance of the invariant presented in section 6.1, readapted to the case where the initial ranking vector is not null. For completeness, let me sketch the proof again. If $\alpha_1^t = 1$, then the inequality (246a) trivially holds. Otherwise, the proof of the inequality (246a) consists of the chain of implications in (247), where I have reasoned exactly as in the case of the chain of implications in (217). Analogously, if $\alpha_2^t = 1$, then the inequality (246b) trivially holds. Otherwise, the proof of the inequality (246b) consists of the two chains of implications in (248) and (249), where again I have reasoned exactly as in the case of the chain of implications in (217).

$$(247) \quad \begin{aligned} \alpha_1^t \geq 1 & \implies \exists s < t \text{ s.t. } \alpha_1^s = \alpha_1^t - 1 \text{ and } \theta^s \text{ not OT-compatible with } \mathbf{a}_1 \\ & \implies \exists s < t \text{ s.t. } \alpha_1^s = \alpha_1^t - 1 \text{ and } \theta_{F_1}^s \leq \theta_{M_1}^s \\ & \implies \exists s < t \text{ s.t. } \alpha_1^s = \alpha_1^t - 1 \text{ and } \alpha_1^s \leq -\alpha_1^s + \alpha_2^s + \theta^{\text{init}} \\ & \implies \exists s < t \text{ s.t. } \alpha_1^s = \alpha_1^t - 1 \text{ and } 2\alpha_1^s \leq \alpha_2^s + \theta^{\text{init}} \\ & \implies 2(\alpha_1^t - 1) \leq \alpha_2^t + \theta^{\text{init}} \\ & \implies \alpha_1^t \leq 1 + \frac{1}{2}(\alpha_2^t + \theta^{\text{init}}) \end{aligned}$$

$$(248) \quad \begin{aligned} \alpha_2^t \geq 1 & \implies \exists s < t \text{ s.t. } \alpha_2^s = \alpha_2^t - 1 \text{ and } \theta^s \text{ not OT-compatible with } \mathbf{a}_2 \\ & \implies \exists s < t \text{ s.t. } \alpha_2^s = \alpha_2^t - 1 \text{ and } \theta_{F_2}^s \leq \theta_{M_2}^s \\ & \implies \exists s < t \text{ s.t. } \alpha_2^s = \alpha_2^t - 1 \text{ and } \alpha_2^s \leq -2\alpha_2 + \theta^{\text{init}} \\ & \implies \exists s < t \text{ s.t. } \alpha_2^s = \alpha_2^t - 1 \text{ and } 3\alpha_2^s \leq \theta^{\text{init}} \\ & \implies 3(\alpha_2^t - 1) \leq \theta^{\text{init}} \\ & \implies \alpha_2^t \leq 1 + \frac{\theta^{\text{init}}}{3} \end{aligned}$$

$$(249) \quad \begin{aligned} \alpha_2^t \geq 1 & \implies \exists s < t \text{ s.t. } \alpha_2^s = \alpha_2^t - 1 \text{ and } \theta^s \text{ not OT-compatible with } \mathbf{a}_2 \\ & \implies \exists s < t \text{ s.t. } \alpha_2^s = \alpha_2^t - 1 \text{ and } \theta_{M_1}^s \leq \theta_{M_2}^s \\ & \implies \exists s < t \text{ s.t. } \alpha_2^s = \alpha_2^t - 1 \text{ and } -\alpha_1 + \alpha_2 + \theta^{\text{init}} \leq -2\alpha_2 + \theta^{\text{init}} \\ & \implies \exists s < t \text{ s.t. } \alpha_2^s = \alpha_2^t - 1 \text{ and } 3\alpha_2^s \leq \alpha_1^s \\ & \implies 3(\alpha_2^t - 1) \leq \alpha_1^t \\ & \implies \alpha_2^t \leq 1 + \frac{1}{3}\alpha_1^t \end{aligned}$$

In order to finally derive the strict inequality (245) from the two inequalities (246), let me distinguish two cases: either $\alpha_1^t < \theta^{\text{init}} - 3$ or else $\alpha_1^t \geq \theta^{\text{init}} - 3$. In the former case $\alpha_1^t < \theta^{\text{init}} - 3$, the strict inequality (245) always holds, as shown by the chain of inequalities in (250). Here, I have reasoned as follows: in step (a), I have used the invariant (246b); in step (b), I have used the fact that we are considering the case $\alpha_1^t < \theta^{\text{init}} - 3$.

$$(250) \quad \alpha_2^t \stackrel{(a)}{\leq} 1 + \frac{1}{3}\alpha_1^t \stackrel{(b)}{<} 1 + \frac{1}{3}(\theta^{\text{init}} - 3) = \frac{1}{3}\theta^{\text{init}}$$

To conclude the proof, I thus only need to show that the opposite case $\alpha_1^t \geq \theta^{\text{init}} - 3$ can never arise. Let me start by noting that the case $\alpha_1^t \geq \theta^{\text{init}}$ indeed can never arise. In fact, in this case the inequality (246b) would become (251a); by replacing (251a) in (246a), we get (251b); the conclusion $\alpha_1^t \leq \frac{2}{3} + \frac{2}{3}\theta^{\text{init}}$ contradicts the hypothesis that we are in a case where $\theta^{\text{init}} \leq \alpha_1^t$ (provided that θ^{init} is large enough that the additive term $\frac{2}{3}$ can be ignored).

$$(251) \quad \begin{aligned} \text{a. } & \alpha_2^t \leq 1 + \frac{1}{3}\theta^{\text{init}}. \\ \text{b. } & \alpha_1^t \leq 1 + \frac{1}{2} \left(\left(1 + \frac{1}{3}\theta^{\text{init}} \right) + \theta^{\text{init}} \right) = \frac{3}{2} + \frac{2}{3}\theta^{\text{init}}. \end{aligned}$$

By reasoning in the same way, one immediately sees that also the three remaining cases $\theta^{\text{init}} = \alpha_1^t - 3$, $\theta^{\text{init}} = \alpha_1^t - 2$ and $\theta^{\text{init}} = \alpha_1^t - 1$ can never arise. ■

Let me make explicit the intuitive idea behind the preceding proof. Consider an arbitrary promotion-demotion update rule (77). Let ν be the amount by which this rule promotes F_2 and M_1 in response to a failure in accounting for the second row of the input comparative tableau (240); let λ be the amount by which this rule demotes M_2 . By reasoning as in the first part of the preceding proof, we conclude that the on-line algorithm (??) with this update rule ends up in the right final vector iff the inequality (252) holds, that generalizes the inequality (245) obtained in the preceding proof for the specific case of the promotion-demotion update rule (238). This inequality (252) says that α_2 must remain small, namely that the second row of the input comparative tableau cannot trigger too many updates. How can we get this effect? The answer was already sketched in the comments above the update rule (123) in section 5.1: if we make λ fairly large, then one exposure to the second row of the input tableau will have the effect of pulling apart the w corresponding to M_1 from the L corresponding to M_2 by a large amount; since that amount is large, it will take a while for it to be eroded; thus, it will take a while for the current ranking vector to become again not OT-incompatible with the second row; in conclusion, the second row will trigger few updates no matter how the rows are fed to the algorithm.

$$(252) \quad (\nu + \lambda)\alpha_2 < \theta^{\text{init}}$$

The promotion-demotion update rule (238) considered here crucially meets this desideratum, because it uses a large value of λ , namely $\lambda = 2$. This explanation immediately predicts that, if we make λ smaller, then the effect of updating the current ranking by the second row of the input tableau will not last long enough, namely the second row will trigger too many updates, namely condition (252) will not hold and the algorithm will end up with the wrong final vector. This prediction is indeed borne out: consider for example the case $\lambda = 1$, so that we in effect have Boersma's (1997) promotion-demotion update rule (79). The on-line algorithm (205) with this update rule might indeed end up with the wrong ranking vector, as shown by the counterexample in (253) and the choice $\theta^{\text{init}} = 7$.

$$(253) \quad \begin{bmatrix} 0 \\ 0 \\ 7 \end{bmatrix} \xrightarrow{\mathbf{a}_2} \begin{bmatrix} 0 \\ 1 \\ 8 \\ 6 \end{bmatrix} \xrightarrow{\mathbf{a}_1} \begin{bmatrix} 1 \\ 1 \\ 7 \\ 6 \end{bmatrix} \xrightarrow{\mathbf{a}_1} \begin{bmatrix} 2 \\ 1 \\ 6 \\ 6 \end{bmatrix} \xrightarrow{\mathbf{a}_1} \begin{bmatrix} 3 \\ 1 \\ 5 \\ 6 \end{bmatrix} \xrightarrow{\mathbf{a}_1} \begin{bmatrix} 4 \\ 1 \\ 4 \\ 6 \end{bmatrix} \xrightarrow{\mathbf{a}_2} \begin{bmatrix} 4 \\ 2 \\ 5 \\ 5 \end{bmatrix} \xrightarrow{\mathbf{a}_2} \begin{bmatrix} 4 \\ 3 \\ 6 \\ 4 \end{bmatrix} \xrightarrow{\mathbf{a}_1} \begin{bmatrix} 5 \\ 3 \\ 5 \\ 4 \end{bmatrix} \xrightarrow{\mathbf{a}_1} \begin{bmatrix} 6 \\ 3 \\ 4 \\ 4 \end{bmatrix} \xrightarrow{\mathbf{a}_2} \begin{bmatrix} 6 \\ 4 \\ 5 \\ 3 \end{bmatrix}$$

Note that in the case of the counterexample (253), the second row triggers a total of four updates before the algorithm converges. This is too much, since $\alpha_2 = 4$ does not satisfy the corresponding inequality (252), with $\nu = \lambda = 1$ and $\theta^{\text{init}} = 7$. The case of the update rule (238) considered here is very different. The on-line algorithm (240) with this promotion-demotion update rule and $\theta^{\text{init}} = 7$ admits only 13 learning paths, listed in (254). And in all 13 paths, the second row of the input comparative tableau triggers only two updates. Indeed the value $\alpha_2 = 2$ does satisfy the inequality (252), with $\nu = 1$, $\lambda = 2$ and $\theta^{\text{init}} = 7$.

$$(254) \quad \begin{array}{cccccccc} \begin{bmatrix} 0 \\ 0 \\ 7 \\ 7 \end{bmatrix} & \xrightarrow{\mathbf{a}_1} & \begin{bmatrix} 1 \\ 0 \\ 6 \\ 7 \end{bmatrix} & \xrightarrow{\mathbf{a}_1} & \begin{bmatrix} 2 \\ 0 \\ 5 \\ 7 \end{bmatrix} & \xrightarrow{\mathbf{a}_1} & \begin{bmatrix} 3 \\ 0 \\ 4 \\ 7 \end{bmatrix} & \xrightarrow{\mathbf{a}_1} & \begin{bmatrix} 4 \\ 0 \\ 3 \\ 7 \end{bmatrix} & \xrightarrow{\mathbf{a}_2} & \begin{bmatrix} 4 \\ 1 \\ 4 \\ 5 \end{bmatrix} & \xrightarrow{\mathbf{a}_2} & \begin{bmatrix} 4 \\ 2 \\ 5 \\ 3 \end{bmatrix} & \xrightarrow{\mathbf{a}_1} & \begin{bmatrix} 5 \\ 2 \\ 4 \\ 3 \end{bmatrix} \\ & & & & & & & & & \xrightarrow{\mathbf{a}_2} & \begin{bmatrix} 3 \\ 1 \\ 5 \\ 5 \end{bmatrix} & \xrightarrow{\mathbf{a}_2} & \begin{bmatrix} 3 \\ 2 \\ 6 \\ 3 \end{bmatrix} & \xrightarrow{\mathbf{a}_2} & \begin{bmatrix} 4 \\ 2 \\ 5 \\ 3 \end{bmatrix} & \xrightarrow{\mathbf{a}_1} & \begin{bmatrix} 5 \\ 2 \\ 4 \\ 3 \end{bmatrix} \\ & & & & & & & & & & \xrightarrow{\mathbf{a}_1} & \begin{bmatrix} 4 \\ 1 \\ 4 \\ 5 \end{bmatrix} & \xrightarrow{\mathbf{a}_1} & \begin{bmatrix} 5 \\ 1 \\ 3 \\ 5 \end{bmatrix} & \xrightarrow{\mathbf{a}_2} & \begin{bmatrix} 5 \\ 2 \\ 4 \\ 3 \end{bmatrix} \end{array}$$

- (255) a. $\mathcal{X} = \mathcal{Y} = \left\{ \begin{array}{cccc} \text{ta,} & \text{da,} & \text{t}^h\text{a,} & \text{d}^h\text{a,} \\ \text{at,} & \text{ad,} & \text{at}^h, & \text{ad}^h, \\ \text{ata,} & \text{ada,} & \text{at}^h\text{a,} & \text{ad}^h\text{a} \end{array} \right\}$
- b. $F_1 = \text{IDENT}[\text{ASPIRATION}][_V]$ $M_1 = *[-\text{SONORANT}, +\text{VOICE}]$
 $F_2 = \text{IDENT}[\text{ASPIRATION}]$ $M_2 = *[\text{+VOICE}][-\text{VOICE}][\text{+VOICE}]$
 $F_3 = \text{IDENT}[\text{VOICE}][_V]$ $M_3 = *[\text{+SPREAD GLOTTIS}]$
 $F_4 = \text{IDENT}[\text{VOICE}]$ $M_4 = *[\text{+SPREAD GLOTTIS}, \text{+VOICE}]$
- c. $\text{Gen}(\text{ta}) = \text{Gen}(\text{da}) = \text{Gen}(\text{t}^h\text{a}) = \text{Gen}(\text{d}^h\text{a}) = \{\text{ta, da, t}^h\text{a, d}^h\text{a}\},$
 $\text{Gen}(\text{at}) = \text{Gen}(\text{ad}) = \text{Gen}(\text{at}^h) = \text{Gen}(\text{ad}^h) = \{\text{at, ad, at}^h, \text{ad}^h\},$
 $\text{Gen}(\text{ada}) = \text{Gen}(\text{ata}) = \text{Gen}(\text{at}^h\text{a}) = \text{Gen}(\text{ad}^h\text{a}) = \{\text{ada, ata, ad}^h\text{a, at}^h\text{a}\}$

Consider the target (whole) language \mathcal{L} in (256). This language corresponds to voicing and aspiration in Korean: voicing is allophonic, namely voicing is allowed only for unaspirated stops in intervocalic position; aspiration is contrastive in prevocalic position but allophonic elsewhere, namely only unaspirated stops are allowed elsewhere.

- (256) $\mathcal{L} = \{\text{ta, at, t}^h\text{a, ada, at}^h\text{a}\}$

Let me construct the phonotactics input tableau corresponding to the target language (256). I start by pairing up the language (256) with the set of underlying/winner/loser form triplets in (257). The second element of each triplet represents the winner surface form, and is taken from the set \mathcal{L} in (256); the first element represents the corresponding underlying form, which is identical to the winner surface form by assumption (34); the third element of the triplet represents a loser surface form; for each pair of an underlying form with the corresponding identical winner surface form, all candidate loser surface forms are considered, displayed on the same row.

- (257) $\left\{ \begin{array}{ccc} (\text{ta, ta, da}), & (\text{ta, ta, t}^h\text{a}), & (\text{ta, ta, d}^h\text{a}), \\ (\text{at, at, ad}), & (\text{at, at, at}^h), & (\text{at, at, ad}^h), \\ (\text{t}^h\text{a, t}^h\text{a, ta}), & (\text{t}^h\text{a, t}^h\text{a, da}), & (\text{t}^h\text{a, t}^h\text{a, d}^h\text{a}), \\ (\text{ada, ada, ata}), & (\text{ada, ada, at}^h\text{a}), & (\text{ada, ada, ad}^h\text{a}), \\ (\text{at}^h\text{a, at}^h\text{a, ata}), & (\text{at}^h\text{a, at}^h\text{a, ada}), & (\text{at}^h\text{a, at}^h\text{a, ad}^h\text{a}) \end{array} \right\}$

Next, I construct the comparative tableau corresponding to the set of triplets in (257), as in (258). Of course, the columns corresponding to the faithfulness constraints contain no single L. In (258), I have ordered the rows by underlying/winner form in blocks of the form $\mathbf{A}_{(x,x,\cdot)}$, with the various blocks corresponding to various forms $x \in \mathcal{L}$ separated by a horizontal line. As usual, I am omitting E's for readability.

	F_1	F_2	F_3	F_4	M_1	M_2	M_3	M_4
(ta, ta, da)			W	W	W			
(ta, ta, t ^h a)	W	W					W	
(ta, ta, d ^h a)	W	W	W	W	W		W	W
(at, at, ad)				W	W			
(at, at, at ^h)		W					W	
(at, at, ad ^h)		W		W	W		W	W
(t ^h a, t ^h a, ta)	W	W					L	
(t ^h a, t ^h a, da)	W	W	W	W	W		L	
(t ^h a, t ^h a, d ^h a)			W	W	W			W
(ada, ada, ata)			W	W	L	W		
(ada, ada, at ^h a)	W	W	W	W	L	W	W	
(ada, ada, ad ^h a)	W	W					W	W
(at ^h a, at ^h a, ata)	W	W					L	
(at ^h a, at ^h a, ada)	W	W	W	W	W	L	L	
(at ^h a, at ^h a, ad ^h a)			W	W	W	L		W

The rows in the comparative tableau (258) that do not contain a single L do not contribute to OT-compatibility, and can therefore be dropped. Furthermore, the row corresponding to (ada, ada, at^ha)

can be ignored, because it is entailed by the row corresponding to (ada, ada, ata). The row corresponding to ($t^h a$, $t^h a$, da) can be ignored, because it is entailed by the row corresponding to ($t^h a$, $t^h a$, ta). Finally, the row corresponding to ($at^h a$, $at^h a$, ata) can be ignored, because it is identical to the row corresponding to ($t^h a$, $t^h a$, ta). In conclusion, we are left with the smaller OT-equivalent comparative tableau (259), with only four comparative rows.

$$(259) \quad \begin{array}{l} (at^h a, at^h a, ada) \\ (t^h a, t^h a, ta) \\ (ada, ada, ata) \\ (at^h a, at^h a, ad^h a) \end{array} \left[\begin{array}{cccccccc} F_1 & F_2 & F_3 & F_4 & M_1 & M_2 & M_3 & M_4 \\ W & W & W & W & W & L & L & \\ W & W & & & & & L & \\ & & W & W & L & W & & \\ & & W & W & W & L & & W \end{array} \right]$$

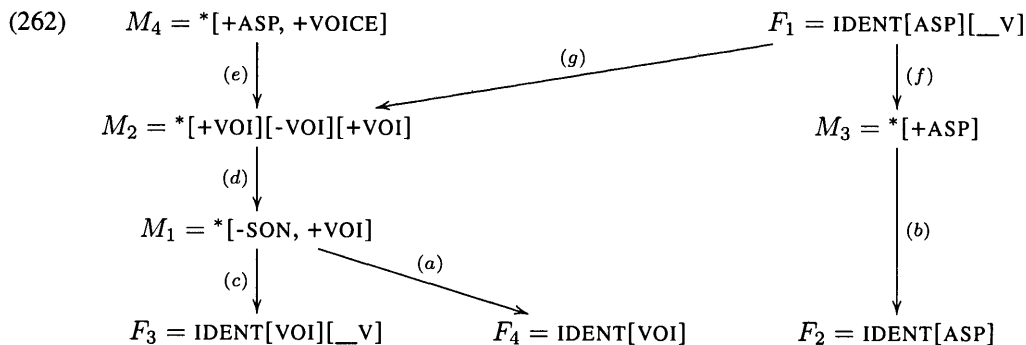
Next, let me construct the sub-tableau of the comparative tableau (258) that only contains the rows corresponding to triplets (x, x, y) such that the loser form y does not belong to the target language (256), as in (260). Here, I have ordered the rows by loser form in blocks of the form $A_{(\cdot, \cdot, y)}$, with the various blocks corresponding to different loser forms $y \in \mathcal{Y} \setminus \mathcal{L}$ separated by a horizontal line.

$$(260) \quad \begin{array}{l} (at, at, ad) \\ (at, at, at^h) \\ (at, at, ad^h) \\ (ta, ta, d^h a) \\ (t^h a, t^h a, d^h a) \\ (ta, ta, da) \\ (t^h a, t^h a, da) \\ (ada, ada, ata) \\ (at^h a, at^h a, ata) \\ (ada, ada, ad^h a) \\ (at^h a, at^h a, ad^h a) \end{array} \left[\begin{array}{cccccccc} F_1 & F_2 & F_3 & F_4 & M_1 & M_2 & M_3 & M_4 \\ & & & W & W & & & \\ \hline & W & & & & & W & \\ \hline & W & & W & W & & W & W \\ \hline W & W & W & W & W & & W & W \\ \hline & & W & W & W & & & \\ \hline W & W & W & W & W & & L & \\ \hline & & W & W & L & W & & \\ \hline W & W & & & & & L & \\ \hline W & W & & & & & W & W \\ \hline & & W & W & W & L & & W \end{array} \right]$$

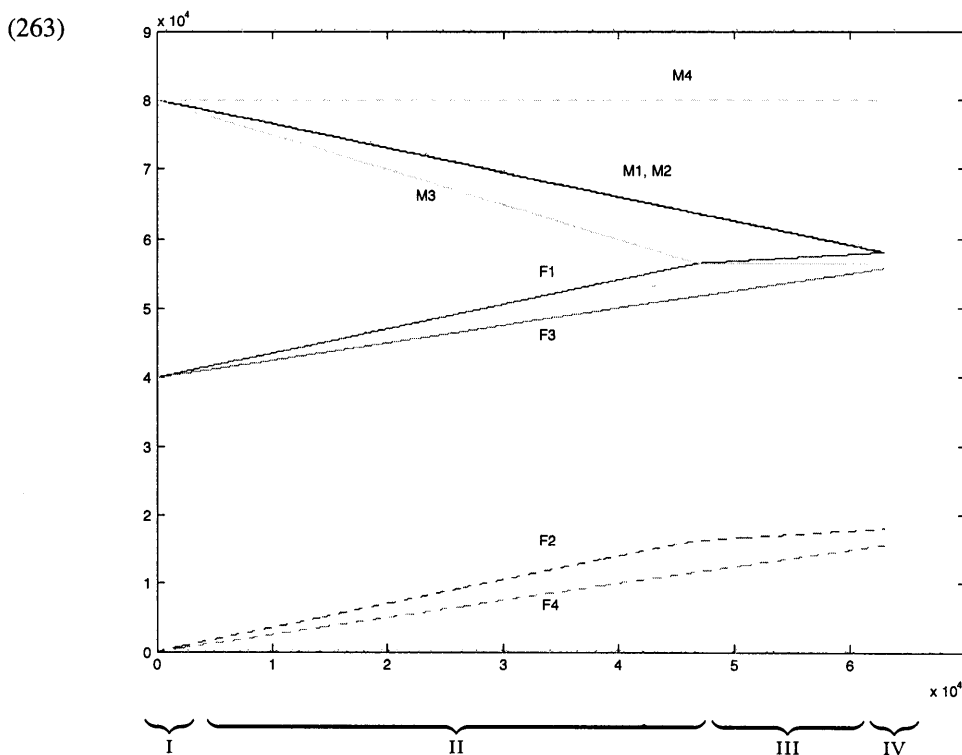
I am now ready to characterize the ranking vectors that OT-correspond to the phonotactics of the language (256), namely such that $\mathcal{L} = \mathcal{R}(\text{OT}_{\gg})$. All four assumptions (224) from subsection ?? are satisfied in the case of the universal specifications (255). Thus, the target rankings \gg can be characterized by means of the two properties (236), that in our case boil down to (261).

- (261) a. \gg is OT-compatible with every row of the comparative tableaux (259);
 b. for every block $A_{(\cdot, \cdot, y)}$ in the comparative tableau (260), there exists at least one row whose \gg -decisive constraint is a markedness constraint.

In order for condition (261b) to hold for the block $A_{(\cdot, \cdot, ad)}$ in the tableau (260), the ranking (262a) must hold. In order for condition (261b) to hold for the block $A_{(\cdot, \cdot, at^h)}$ in the tableau (260), the ranking (262b) must hold. In order for condition (261b) to hold for the block $A_{(\cdot, \cdot, da)}$ in the tableau (260), the ranking (262c) must hold. In order for condition (261a) to hold for the row (ada, ada, ata) in the comparative tableau (259) given the two rankings (262a) and (262c) established so far, the ranking (262d) must hold. Condition (261b) is then satisfied for the block $A_{(\cdot, \cdot, ata)}$ in the tableau (260). In order for condition (261a) to hold for the row (at^h , $at^h a$, $ad^h a$) of the comparative tableau (259) given the rankings (262a), (262c) and (262d) established so far, the ranking (262e) must hold. Condition (261b) for the blocks $A_{(\cdot, \cdot, adha)}$, $A_{(\cdot, \cdot, dha)}$ and $A_{(\cdot, \cdot, ad^h)}$ in the comparative tableau (260) is satisfied by the rankings already established. In order for condition (261a) to hold for the row ($t^h a$, tha , ta) in the comparative tableau (259) given the ranking (262b) already established, the ranking (262f) must hold. Finally, in order for condition (261a) to hold for the row ($at^h a$, $at^h a$, ada) in the comparative tableau (259) given the rankings already established, the ranking (262g) must hold. In conclusion, a ranking vector



I have run the on-line algorithm (205) with the promotion-demotion update rule (238) starting from the biased initial vector (237b) with $\theta^{\text{init}} = 40,000$ on the input comparative tableau (259) with the rows sampled uniformly. The dynamics of the ranking values in one such run is reported in (263), and the corresponding final ranking vector is provided in (264), which correctly accounts for all the rankings in (262). The reason for such a high value of the constant θ^{init} is that it smoothes down the randomness in the choice of the rows, thus yielding the apparently straight lines in (263).



(264) $\theta = [\begin{array}{cccccccc} F_1 & F_2 & F_3 & F_4 & M_1 & M_2 & M_3 & M_4 \\ 58043 & 18043 & 55685 & 15685 & 58041 & 58042 & 56460 & 80001 \end{array}]$

In the rest of this section, I study in detail with paper and pencil the behavior of the algorithm, thus explain the diagram in (263).

6.3.2.1 Stage I

At the beginning of the learning process, we have all markedness constraints ranked above all faithfulness constraints. After at most one exposure to the fourth row of the comparative tableau (259), that row will be accounted for by the decisive constraint M_4 . And any subsequent ranking vector entertained by the algorithm will still account for that row, since the constraint M_4 will not move further while the constraint M_2 will keep dropping. Let T_1 be the time at which M_4 crosses M_2

and thus stage I ends. Under the assumption that the rows of the input comparative tableau (259) are sampled uniformly, stage I is over very soon,² namely T_I is very small. If θ^{init} is rather large, I can safely assume that no substantial modification has happened to the initial ranking of the other various constraints. I will thus assume that the ranking vector θ^{T_I} entertained at the end of stage I is identical to the initial ranking vector θ^{init} , as stated in (265).

$$(265) \quad \begin{aligned} \theta_{M_1}^{T_I} = \theta_{M_2}^{T_I} = \theta_{M_3}^{T_I} &= 2\theta^{\text{init}} \\ \theta_{F_1}^{T_I} = \theta_{F_3}^{T_I} &= \theta^{\text{init}} \\ \theta_{F_2}^{T_I} = \theta_{F_4}^{T_I} &= 0 \end{aligned}$$

Note indeed that stage I is so short in the simulation described in (263) that it is not even visible in the diagram.

6.3.2.2 Stage II

From now on, the bottom row of the comparative tableau (259) can be ignored, since it won't trigger any further update. And constraint M_4 can be ignored as well, since its ranking value won't change. The initial ranking vector of stage II is the final ranking vector (265) of stage I. Thus, to study stage II means to study the behavior of the algorithm with the initial ranking vector (265) and the input tableau (266), obtained from (259) by deleting the last columns and the last row.

$$(266) \quad \begin{array}{l} \text{row 1} \\ \text{row 2} \\ \text{row 3} \end{array} \begin{bmatrix} F_1 & F_2 & F_3 & F_4 & M_1 & M_2 & M_3 \\ W & W & W & W & W & L & L \\ W & W & & & & & L \\ & & W & W & L & W & \end{bmatrix}$$

Let α_1^t , α_2^t and α_3^t be the dual variables corresponding to the three rows of the comparative tableau (266) at a generic time t of stage II. More explicitly, α_i^t is the number of times that the algorithm has updated its current ranking vector in response to a failure in accounting for the i th row of the tableau (266) among the first t updates performed within stage II.

First part of the analysis The analysis of the behavior of the algorithm in stage II rests on the two approximate relations (267) between the two dual variables α_1^t and α_3^t at a generic time t in stage II. Let me explain intuitively why these two relations hold.

$$(267) \quad \begin{aligned} \text{a. } \alpha_3^t &\approx \frac{3}{2}\alpha_1^t \\ \text{b. } \alpha_1^t + \alpha_3^t &\approx \frac{1}{4}t \end{aligned}$$

The approximate relation (267a) immediately follows from the two inequalities in (268), since the two different terms $-3/2$ and $+1$ at the left and at the right of (268) can be ignored when α_1^t and α_3^t are large.

$$(268) \quad \frac{3}{2}\alpha_1^t - \frac{3}{2} \leq \alpha_3^t \leq \frac{3}{2}\alpha_1^t + 1$$

Let me prove the right hand side inequality in (268). If $\alpha_3^t = 0$, then this inequality trivially holds. Otherwise, the proof of this inequality consists of the chain of implication in (269), where I have reasoned exactly as in the case of the chain of implications in (217).

²If they are not sampled uniformly, then it might take a very long time for the fourth row to be fed to the algorithm and thus stage I might in principle take very long.

$$\begin{aligned}
(269) \quad \alpha_3^t \geq 1 &\implies \exists s < t \text{ s.t. } \alpha_3^s = \alpha_3^t - 1 \text{ and } \theta^s \text{ not OT-compatible with the third row} \\
&\implies \exists s < t \text{ s.t. } \alpha_3^s = \alpha_3^t - 1 \text{ and } \theta_{M_2}^s \leq \theta_{M_1}^s \\
&\implies \exists s < t \text{ s.t. } \alpha_3^s = \alpha_3^t - 1 \text{ and } -5\alpha_1^s + \alpha_3^s \leq \alpha_1^s - 3\alpha_3^s \\
&\implies \exists s < t \text{ s.t. } \alpha_3^s = \alpha_3^t - 1 \text{ and } 4\alpha_3^s \leq 6\alpha_1^s \\
&\implies \exists s < t \text{ s.t. } \alpha_3^s = \alpha_3^t - 1 \text{ and } 4\alpha_3^s \leq 6\alpha_1^t \\
&\implies \exists s < t \text{ s.t. } \alpha_3^s = \alpha_3^t - 1 \text{ and } 4(\alpha_3^t - 1) \leq 6\alpha_1^t \\
&\implies \exists s < t \text{ s.t. } \alpha_3^s = \alpha_3^t - 1 \text{ and } \alpha_3^t \leq \frac{3}{2}\alpha_1^t + 1
\end{aligned}$$

The proof of the left hand side inequality in (268) is identical; I provide it explicitly for completeness. If $\alpha_1^t = 0$, then this inequality trivially holds. Otherwise, the proof of this inequality consists of the chain of implication in (270), where I have reasoned once again exactly as in the case of the chain of implications in (217).

$$\begin{aligned}
(270) \quad \alpha_1^t \geq 1 &\implies \exists s < t \text{ s.t. } \alpha_1^s = \alpha_1^t - 1 \text{ and } \theta^s \text{ not OT-compatible with the first row} \\
&\implies \exists s < t \text{ s.t. } \alpha_3^s = \alpha_3^t - 1 \text{ and } \theta_{M_2}^s \geq \theta_{M_1}^s \\
&\implies \exists s < t \text{ s.t. } \alpha_3^s = \alpha_3^t - 1 \text{ and } -5\alpha_1^s + \alpha_3^s \geq \alpha_1^s - 3\alpha_3^s \\
&\implies \exists s < t \text{ s.t. } \alpha_3^s = \alpha_3^t - 1 \text{ and } 4\alpha_3^s \geq 6\alpha_1^s \\
&\implies \exists s < t \text{ s.t. } \alpha_3^s = \alpha_3^t - 1 \text{ and } 4\alpha_3^t \geq 6\alpha_1^s \\
&\implies \exists s < t \text{ s.t. } \alpha_3^s = \alpha_3^t - 1 \text{ and } 4\alpha_3^t \geq 6(\alpha_1^t - 1) \\
&\implies \exists s < t \text{ s.t. } \alpha_3^s = \alpha_3^t - 1 \text{ and } \alpha_3^t \geq \frac{3}{2}\alpha_1^t - \frac{3}{2}
\end{aligned}$$

Let me now turn to the approximate relation (267b). The approximate relation (267a) between the two dual variables α_1^t and α_3^t entails in particular that the ranking values $\theta_{M_1}^t$ and $\theta_{M_2}^t$ of the two constraints M_1 and M_2 are very close to each other throughout stage II, as shown by the two computations in (271).

$$\begin{aligned}
(271) \quad \theta_{M_1}^t &= \alpha_1^t - 3\alpha_3^t + 2\theta^{\text{init}} \approx 2\theta^{\text{init}} - \frac{7}{20}\alpha_1 \\
\theta_{M_2}^t &= -5\alpha_1^t + \alpha_3^t + 2\theta^{\text{init}} \approx 2\theta^{\text{init}} - \frac{7}{20}\alpha_1
\end{aligned}$$

Let Ω_1^t be the event that the current ranking vector at time t is not OT-compatible with the first row and let Ω_3^t be the probability that it is not OT-compatible with the third row. The approximate relation (267b) can be explained as in (272). In step (a), I have noted that each row get sampled approximately 1/4 of the times, since the four rows of the comparative tableau (259) are sampled uniformly. In step (b), I have reasoned as follows: the OT-compatibility of the current ranking vector with either the first or the third row only depends on the relative ranking of M_1 and M_2 , since the constraint M_3 will always be below M_2 and furthermore all faithfulness constraints are still below all markedness constraints throughout stage II. At every time t , the current ranking vector ranks M_1 above M_2 or vice versa, so that it is either OT-compatible with the first row or with the third.

$$\begin{aligned}
(272) \quad \alpha_1^t + \alpha_3^t &\stackrel{(a)}{=} \frac{t}{4}\mathbb{P}[\Omega_1^t] + \frac{t}{4}\mathbb{P}[\Omega_3^t] \\
&= \frac{t}{4}\left(\mathbb{P}[\Omega_1^t] + \mathbb{P}[\Omega_3^t]\right) \\
&\stackrel{(b)}{=} \frac{t}{4}
\end{aligned}$$

Second part of the analysis Given the preliminary relations (267), the three dual variables α_1^t , α_2^t and α_3^t at each time t can be approximately expressed as a function of t as in (273). The two approximate expressions (273a) and (273c) immediately follow from (267). The approximate expression (273b) can be justified as follows: since the four rows of the comparative tableau (259) are sampled uniformly, we expect that at time t the second row will have been sampled approximately $t/4$ times; each time the row is sampled and fed to the algorithm, it does trigger an update, because the markedness constraint M_3 is above both faithfulness constraints F_1 and F_2 throughout the entire stage II; in conclusion, we thus expect that $\alpha_2^t \approx t/4$.

$$(273) \quad \begin{aligned} \text{a. } \alpha_1^t &\approx \frac{1}{10}t \\ \text{b. } \alpha_2^t &\approx \frac{1}{4}t \\ \text{c. } \alpha_3^t &\approx \frac{3}{20}t \end{aligned}$$

I report in (274) the actual value of the three dual variables entertained by the algorithm in the simulation represented in the diagram (263) at two different times. The values in (274) are in perfect accord with the estimates in (273).

(274)	$t = 20,000$	$t = 40,000$
α_1^t	2,000	4,012
α_2^t	5,030	10,031
α_3^t	3,000	6,019

The approximate expressions (273) for the dual variables at a generic time t in stage II yield the approximate expressions (275) for the primal variables at a generic time t in stage II as a function of t . These computations are trivial, but for a small remark concerning M_3 . The column corresponding to M_3 has two L's, corresponding to the first two rows. The L in the second row will be undominated throughout the entire stage II, and thus contributes to the reranking of constraint M_3 . But the L corresponding to the first row is dominated by the winner preferring constraint M_1 throughout the entire stage II. The reason is as follows: clearly, M_3 will be ranked below M_2 throughout the entire stage II; as noted in (271), M_2 and M_1 have approximately the same ranking throughout the entire stage II; thus, M_3 will be ranked below M_1 throughout the entire stage II; hence, the L corresponding to M_3 in the first row is never undominated and thus never contributes to the reranking of M_3 . The estimates in (275) are in very good accord with the diagram (263).

$$(275) \quad \begin{aligned} \theta_{F_1}^t &= \alpha_1^t + \alpha_2^t + \theta^{\text{init}} = \frac{1}{10}t + \frac{1}{4}t + \theta^{\text{init}} &&= \frac{7}{20}t + \theta^{\text{init}} \\ \theta_{F_2}^t & &&= \frac{7}{20}t \\ \theta_{F_3}^t &= \alpha_1^t + \alpha_3^t + \theta^{\text{init}} = \frac{1}{10}t + \frac{3}{20}t + \theta^{\text{init}} &&= \frac{5}{20}t + \theta^{\text{init}} \\ \theta_{F_4}^t & &&= \frac{5}{20}t \\ \theta_{M_1}^t &= \alpha_1^t - 3\alpha_3^t + 2\theta^{\text{init}} = \frac{1}{10}t - 3\frac{3}{20}t + 2\theta^{\text{init}} &&= 2\theta^{\text{init}} - \frac{7}{20}t \\ \theta_{M_2}^t &= -5\alpha_1^t + \alpha_3^t + 2\theta^{\text{init}} = -5\frac{1}{10}t + \frac{3}{20}t + 2\theta^{\text{init}} &&= 2\theta^{\text{init}} - \frac{7}{20}t \\ \theta_{M_3}^t &= -2\alpha_2^t + 2\theta^{\text{init}} = -2\frac{1}{4}t + 2\theta^{\text{init}} &&= 2\theta^{\text{init}} - \frac{1}{2}t \end{aligned}$$

From the estimates in (275), we immediately see that at every time t within stage II, we have that: the largest ranking value among faithfulness constraints is that of F_1 ; and furthermore that the smallest ranking value among markedness constraints is that of M_3 . Thus, stage II ends when F_1 crosses M_3 and the second row is thus correctly accounted for by the decisive F_1 . Let T_{II} be the time at which this crossing of M_3 by F_1 happens, and stage II thus ends. This time T_{II} can be easily determined as a function of θ^{init} as in (276). Note that for the case $\theta^{\text{init}} = 40,000$, the estimate $T_{\text{II}} \approx \frac{20}{17}\theta^{\text{init}}$ derived in (276) yields $T_{\text{II}} \approx 47,059$, which accords perfectly with the diagram (263).

$$(276) \quad \begin{aligned} \theta_{F_1}^{T_{\text{II}}} = \theta_{M_3}^{T_{\text{II}}} &\iff \frac{7}{20}T_{\text{II}} + \theta^{\text{init}} = 2\theta^{\text{init}} - \frac{1}{2}T_{\text{II}} \\ &\iff T_{\text{II}} \approx \frac{20}{17}\theta^{\text{init}} \end{aligned}$$

From the general estimates (275) of the components of the ranking vector during stage II together with the estimate (276) of the final time T_{II} of stage II, we get the characterization (277) of the ranking vector $\theta^{T_{\text{II}}}$ entertained by the algorithm at the end of stage II.

$$\begin{aligned}
(277) \quad \theta_1^{T_{II}} &= \frac{7}{20} \frac{20}{17} \theta^{\text{init}} + \theta^{\text{init}} = \frac{24}{17} \theta^{\text{init}} \\
\theta_2^{T_{II}} &= \frac{7}{20} \frac{20}{17} \theta^{\text{init}} = \frac{7}{17} \theta^{\text{init}} \\
\theta_3^{T_{II}} &= \frac{5}{20} \frac{20}{17} \theta^{\text{init}} + \theta^{\text{init}} = \frac{22}{17} \theta^{\text{init}} \\
\theta_3^{T_{II}} &= \frac{5}{20} \frac{20}{17} \theta^{\text{init}} = \frac{5}{17} \theta^{\text{init}} \\
\theta_5^{T_{II}} &= 2\theta^{\text{init}} - \frac{7}{20} \frac{20}{17} \theta^{\text{init}} = \frac{27}{17} \theta^{\text{init}} \\
\theta_6^{T_{II}} &= 2\theta^{\text{init}} - \frac{7}{20} \frac{20}{17} \theta^{\text{init}} = \frac{27}{17} \theta^{\text{init}} \\
\theta_7^{T_{II}} &= 2\theta^{\text{init}} - \frac{1}{2} \frac{20}{17} \theta^{\text{init}} = \frac{24}{17} \theta^{\text{init}}
\end{aligned}$$

6.3.2.3 Stage III

From now on, the second row of the comparative tableau (266) can be ignored, since it won't trigger any further update. And constraint M_3 can be ignored as well, since its ranking value won't change. The initial ranking vector of stage III is the final ranking vector (277) of stage II. Thus, to study stage III means to study the behavior of the algorithm with the initial ranking vector (277) and the input tableau (278), obtained from (259) by deleting the last two columns and the second and fourth rows.

$$(278) \quad \begin{array}{c} \text{row 1} \\ \text{row 3} \end{array} \left[\begin{array}{cccccc} F_1 & F_2 & F_3 & F_4 & M_1 & M_2 \\ \text{W} & \text{W} & \text{W} & \text{W} & \text{W} & \text{L} \\ & & \text{W} & \text{W} & \text{L} & \text{W} \end{array} \right]$$

Let α_1^t and α_3^t be the dual variables associated with the first and the third comparative rows at a generic time t in stage III. Of course, the approximate characterization (273a) and (273c) of these two dual variables α_1^t and α_3^t obtained for stage II also holds at stage III. Thus, the primal variables at a generic time t in stage III can be approximately expressed as a function of time t as in (279).

$$\begin{aligned}
(279) \quad \theta_{F_1}^t &= \alpha_1^t + \theta_{F_1}^{T_{II}} = \frac{1}{10}t + \frac{24}{17}\theta^{\text{init}} \\
\theta_{F_2}^t &= \alpha_1^t + \theta_{F_2}^{T_{II}} = \frac{1}{10}t + \frac{7}{17}\theta^{\text{init}} \\
\theta_{F_3}^t &= \alpha_1^t + \alpha_3^t + \theta_{F_3}^{T_{II}} = \frac{5}{20}t + \frac{22}{17}\theta^{\text{init}} \\
\theta_{F_4}^t &= \alpha_1^t + \alpha_3^t + \theta_{F_4}^{T_{II}} = \frac{5}{20}t + \frac{5}{17}\theta^{\text{init}} \\
\theta_{M_1}^t &= \alpha_1^t - 3\alpha_3^t + \theta_{M_1}^{T_{II}} = -\frac{7}{20}t + \frac{27}{17}\theta^{\text{init}} \\
\theta_{M_2}^t &= -5\alpha_1^t + \alpha_3^t + \theta_{M_2}^{T_{II}} = -\frac{7}{20}t + \frac{27}{17}\theta^{\text{init}}
\end{aligned}$$

Stage III ends when a faithfulness constraint crosses the two markedness constraints, that drop approximately at the same speed. Since F_1 is always above F_2 and F_3 is always above F_4 , then the crossing faithfulness constraint must be either F_1 or F_3 . Thus, I need to determine which one is the first crossing constraint. Contrary to the case of stage II, in the case of stage III a qualitative analysis is not sufficient to settle the issue: in fact, (277) shows that F_1 starts out higher than F_3 at the beginning of stage III; yet, (279) shows that F_1 raises slower than F_3 . A quantitative analysis is needed instead. This quantitative analysis is provided in (280).

$$\begin{aligned}
(280) \quad \text{a. } F_1 \text{ crosses } M_1, M_2 &\iff \theta_{F_1}^t = \theta_{M_1}^t \\
&\iff \frac{1}{10}t + \frac{24}{17}\theta^{\text{init}} = -\frac{7}{20}t + \frac{27}{17}\theta^{\text{init}} \\
&\iff t = \frac{20}{51}\theta^{\text{init}} \\
\text{b. } F_3 \text{ crosses } M_1, M_2 &\iff \theta_{F_3}^t = \theta_{M_1}^t \\
&\iff \frac{5}{20}t + \frac{22}{17}\theta^{\text{init}} = -\frac{7}{20}t + \frac{27}{17}\theta^{\text{init}} \\
&\iff t = \frac{25}{51}\theta^{\text{init}}
\end{aligned}$$

Thus, stage III ends when F_1 crosses both M_1 and M_2 , and this happens after $T_{III} = \frac{20}{51}\theta^{\text{init}}$ updates after the beginning of stage III. Thus, the ranking vector $\theta^{T_{III}}$ entertained by the algorithm at the end of stage III is (281).

$$\begin{aligned}
 (281) \quad \theta_{F_1}^{T_{III}} &= \frac{1}{10} \frac{20}{51} \theta^{init} + \frac{24}{17} \theta^{init} = \theta^{init} \\
 \theta_{F_2}^{T_{III}} &= \frac{1}{10} \frac{20}{51} \theta^{init} + \frac{7}{17} \theta^{init} = \theta^{init} \\
 \theta_{F_3}^{T_{III}} &= \frac{5}{20} \frac{20}{51} \theta^{init} + \frac{22}{17} \theta^{init} = \theta^{init} \\
 \theta_{F_4}^{T_{III}} &= \frac{5}{20} \frac{20}{51} \theta^{init} + \frac{5}{17} \theta^{init} = \theta^{init} \\
 \theta_{M_1}^{T_{III}} &= -\frac{7}{20} \frac{20}{51} \theta^{init} + \frac{27}{17} \theta^{init} = \theta^{init} \\
 \theta_{M_2}^{T_{III}} &= -\frac{7}{20} \frac{20}{51} \theta^{init} + \frac{27}{17} \theta^{init} = \theta^{init}
 \end{aligned}$$

6.3.2.4 Stage IV

From now on, the first row of the comparative tableau (278) can be ignored, since it won't trigger any further update. The initial ranking vector of stage IV is the final ranking vector (281) of stage III. Thus, to study stage IV means to study the behavior of the algorithm with the initial ranking vector (281) and the input tableau (282), obtained from (259) by deleting all rows but the third one.

$$(282) \quad \text{row 3} \quad \left[\begin{array}{cccccc} F_1 & F_2 & F_3 & F_4 & M_1 & M_2 \\ & & W & W & L & W \end{array} \right]$$

As shown in (281), the ranking values of the two markedness constraints M_1 and M_2 are very close at the beginning of stage IV and far away from the ranking value of F_3 . Thus, a few updates will be sufficient in stage IV in order for M_2 to settle above M_1 , so that the current ranking vector will become OT-compatible with the tableau (282) and learning will cease.

6.3.3 Third test case: the whole Azba typology

In this subsection, I study in detail the case of the Azba typology discussed in Prince and Tesar (2004). Consider the universal specifications $(\mathcal{X}, \mathcal{Y}, Gen, C)$ in (283). These universal specifications are centered around the two features [STOP-VOICING] and [FRICATIVE-VOICING]. Each one of the two features is associated with a set of a general faithfulness constraint, a positional one and a markedness one; the markedness constraint $M_1 = AGREE$, which requires adjacent obstruents to agree in voicing, links the two sets of constraints together.

$$\begin{aligned}
 (283) \quad \text{a. } \mathcal{X} = \mathcal{Y} &= \left\{ \begin{array}{cccc} \text{pa,} & \text{ba,} & \text{ap,} & \text{ab,} \\ \text{sa,} & \text{za,} & \text{as,} & \text{az,} \\ \text{apsa,} & \text{apza,} & \text{absa,} & \text{abza,} \\ \text{aspa,} & \text{azpa,} & \text{asba,} & \text{azba} \end{array} \right\} \\
 \text{b. } \left. \begin{array}{l} Gen(\text{pa}) \\ Gen(\text{ba}) \end{array} \right\} &= \{\text{pa, ba}\} & \left. \begin{array}{l} Gen(\text{ap}) \\ Gen(\text{ab}) \end{array} \right\} &= \{\text{ap, ab}\} \\
 \left. \begin{array}{l} Gen(\text{sa}) \\ Gen(\text{za}) \end{array} \right\} &= \{\text{sa, za}\} & \left. \begin{array}{l} Gen(\text{as}) \\ Gen(\text{az}) \end{array} \right\} &= \{\text{as, az}\} \\
 \left. \begin{array}{l} Gen(\text{apsa}) \\ Gen(\text{apza}) \\ Gen(\text{absa}) \\ Gen(\text{abza}) \end{array} \right\} &= \left\{ \begin{array}{l} \text{apsa, apza,} \\ \text{absa, abza} \end{array} \right\} & \left. \begin{array}{l} Gen(\text{aspa}) \\ Gen(\text{azpa}) \\ Gen(\text{asba}) \\ Gen(\text{azba}) \end{array} \right\} &= \left\{ \begin{array}{l} \text{apsa, azpa,} \\ \text{asba, azba} \end{array} \right\} \\
 \text{c. } C &= \left\{ \begin{array}{l} F_1 = \text{IDENT[STOP-VOICING]} \\ F_2 = \text{IDENT[FRICATIVE-VOICING]} \\ F_3 = \text{IDENT[STOP-VOICING][ONSET]} \\ F_4 = \text{IDENT[FRICATIVE-VOICING][ONSET]} \\ M_1 = \text{AGREE[VOICE]} \\ M_2 = *[\text{+STOP-VOICING}] \\ M_3 = *[\text{+FRICATIVE-VOICING}] \end{array} \right\}
 \end{aligned}$$

There are a total of 41 grammars in the typology described by the universal specifications (283), as computed by the software OTSoft by Hayes et al. (2003). These 41 grammars correspond to a total of

37 different languages (a given language might correspond to two different grammars, since different grammars might neutralize a given form in different ways). I have listed all these languages in (284), omitting the two trivial languages, namely the maximal one corresponding to the entire set of surface forms and the minimal one corresponding only to the unmarked forms {pa, ap, sa, as, apsa, aspa}.

(284)

I.	$\left\{ \begin{array}{l} \text{pa, ba, ap, ab,} \\ \text{sa, za, as, az,} \\ \text{apsa, apza, absa, abza,} \\ \text{aspa, azba} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{pa, ba, ap, ab,} \\ \text{sa, za, as, az,} \\ \text{apsa, abza,} \\ \text{aspa, azpa, asba, azba} \end{array} \right\}$
II.	$\left\{ \begin{array}{l} \text{pa, ba, ap, ab,} \\ \text{sa, za, as, az,} \\ \text{apsa, abza,} \\ \text{aspa, azba} \end{array} \right\}$	
III.	$\left\{ \begin{array}{l} \text{pa, ba, ap, ab,} \\ \text{sa, za, as, abza,} \\ \text{apsa, apza, absa, abza,} \\ \text{aspa, azba} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{pa, ba, ap, az,} \\ \text{sa, za, as, abza,} \\ \text{apsa, abza,} \\ \text{aspa, azpa, asba, azba} \end{array} \right\}$
IV.	$\left\{ \begin{array}{l} \text{pa, ba, ap, ab,} \\ \text{sa, za, as, abza,} \\ \text{apsa, apza, absa, abza,} \\ \text{aspa, asba,} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{pa, ba, ap, az,} \\ \text{sa, za, as, abza,} \\ \text{apsa, apza, asba, azba} \end{array} \right\}$
V.	$\left\{ \begin{array}{l} \text{pa, ba, ap, ab,} \\ \text{sa, za, as, abza,} \\ \text{apsa, azba} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{pa, ba, ap, az,} \\ \text{sa, za, as, abza,} \\ \text{apsa, azba} \end{array} \right\}$
VI.	$\left\{ \begin{array}{l} \text{pa, ba, ap, ab,} \\ \text{sa, za, as, abza,} \\ \text{apsa,} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{pa, ba, ap, az,} \\ \text{sa, za, as, abza,} \\ \text{apsa,} \end{array} \right\}$
VII.	$\left\{ \begin{array}{l} \text{pa, ba, ap, ab,} \\ \text{sa, za, as, abza,} \\ \text{apsa, asba,} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{pa, ba, ap, az,} \\ \text{sa, za, as, abza,} \\ \text{apsa, apza,} \end{array} \right\}$
VIII.	$\left\{ \begin{array}{l} \text{pa, ba, ap, ab,} \\ \text{sa, za, as,} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{pa, ap, az,} \\ \text{sa, za, as,} \end{array} \right\}$
IX.	$\left\{ \begin{array}{l} \text{pa, ba, ap, ab,} \\ \text{sa, za, as, azba} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{pa, ap, az,} \\ \text{sa, za, as, abza,} \end{array} \right\}$
X.	$\left\{ \begin{array}{l} \text{pa, ba, ap, ab,} \\ \text{sa, za, as, asba} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{pa, ap, az,} \\ \text{sa, za, as,} \end{array} \right\}$
XI.	$\left\{ \begin{array}{l} \text{pa, ba, ap, ab,} \\ \text{sa, za, as, abza,} \\ \text{apsa, azba} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{pa, ap, az,} \\ \text{sa, za, as, abza,} \\ \text{apsa, azba} \end{array} \right\}$

XII.	$\left\{ \begin{array}{l} \text{pa, ba, ap, ab,} \\ \text{sa, as,} \\ \text{apsa, absa,} \\ \text{aspa, asba,} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{pa, ap,} \\ \text{sa, za, as, az,} \\ \text{apsa, apza,} \\ \text{aspa, azpa,} \end{array} \right\}$
XIII.	$\left\{ \begin{array}{l} \text{pa, ba, ap,} \\ \text{sa, za, as,} \\ \text{apsa, apza,} \\ \text{aspa, azba} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{pa, ba, ap,} \\ \text{sa, za, as,} \\ \text{apsa, abza,} \\ \text{aspa, asba,} \end{array} \right\}$
XIV.	$\left\{ \begin{array}{l} \text{pa, ba, ap,} \\ \text{sa, za, as,} \\ \text{apsa, apza,} \\ \text{aspa, asba,} \end{array} \right\}$	
XV.	$\left\{ \begin{array}{l} \text{pa, ba, ap,} \\ \text{sa, za, as,} \\ \text{apsa, azba} \\ \text{aspa,} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{pa, ba, ap,} \\ \text{sa, za, as,} \\ \text{apsa, abza,} \\ \text{aspa,} \end{array} \right\}$
XVI.	$\left\{ \begin{array}{l} \text{pa, ba, ap,} \\ \text{sa, za, as,} \\ \text{apsa, abza,} \\ \text{aspa, azba} \end{array} \right\}$	
XVII.	$\left\{ \begin{array}{l} \text{pa, ba, ap,} \\ \text{sa, as,} \\ \text{apsa,} \\ \text{aspa,} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{pa, ap,} \\ \text{sa, za, as,} \\ \text{apsa,} \\ \text{aspa,} \end{array} \right\}$
XVIII.	$\left\{ \begin{array}{l} \text{pa, ba, ap,} \\ \text{sa, as,} \\ \text{apsa, abza,} \\ \text{aspa, azba} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{pa, ap,} \\ \text{sa, za, as,} \\ \text{apsa, abza,} \\ \text{aspa,} \end{array} \right\}$
XIX.	$\left\{ \begin{array}{l} \text{pa, ba, ap,} \\ \text{sa, as,} \\ \text{apsa, asba,} \\ \text{aspa,} \end{array} \right\}$	$\left\{ \begin{array}{l} \text{pa, ap,} \\ \text{sa, za, as,} \\ \text{apsa, apza,} \\ \text{aspa,} \end{array} \right\}$

Each language in the right column of (284) is obtained from the corresponding language in the left column through the correspondences in (285), and vice versa. Since the universal specifications (283) are symmetric w.r.t. the two features [STOP-VOICING] and [FRICATIVE-VOICING], I can then ignore the languages in the right column and only concentrate on the languages in the left column. In conclusion, I am left with 19 test cases.

(285) $\text{ba} \Leftrightarrow \text{za}$ $\text{ab} \Leftrightarrow \text{az}$
 $\text{apza} \Leftrightarrow \text{asba}$ $\text{apsa} \Leftrightarrow \text{aspa}$
 $\text{abza} \Leftrightarrow \text{azba}$ $\text{absa} \Leftrightarrow \text{azpa}$
 $\text{azpa} \Leftrightarrow \text{absa}$ $\text{aspa} \Leftrightarrow \text{apsa}$
 $\text{azba} \Leftrightarrow \text{abza}$ $\text{asba} \Leftrightarrow \text{apza}$

I have run the OT on-line algorithm (205) with the promotion-demotion update rule (238), with the initial ranking of the form (237b) with $\theta^{\text{init}} = 1000$ and with uniform sampling of the rows of the input phonotactic comparative tableau corresponding to each of the 19 languages in the left column in (284). I report in (286) the final ranking vector obtained for each of the 19 languages, together with the unique ranking it represents. Multiple simulations with the same comparative tableaux always yielded the same result (the actual components of the final ranking vector might differ slightly from one simulation to the other because of the randomization in the sampling of the rows, but not the

ranking it represents). It turns out that the final ranking entertained by the model OT-corresponds to the target language in all cases but the case of language XI.

(286)	<i>Final ranking vector</i>	<i>Corresponding ranking</i>
I.	(1000, 857, 1177, 1367, 998, 997, 856)	$F4 \gg F3 \gg F1 \gg M1 \gg M2 \gg F2 \gg M3$
II.	(765, 774, 1174, 1184, 2000, 763, 773)	$M1 \gg F4 \gg F3 \gg F2 \gg M3 \gg F1 \gg M2$
III.	(1019, 567, 1206, 1363, 1018, 1015, 1016)	$F4 \gg F3 \gg F1 \gg M1 \gg M3 \gg M2 \gg F2$
IV.	(990, 424, 1205, 1351, 988, 989, 1349)	$F4 \gg M3 \gg F3 \gg F1 \gg M2 \gg M1 \gg F2$
V.	(787, 305, 1191, 1208, 2001, 786, 1188)	$M1 \gg F4 \gg F3 \gg M3 \gg F1 \gg M2 \gg F2$
VI.	(809, 286, 1161, 1286, 2000, 807, 1285)	$M1 \gg F4 \gg M3 \gg F3 \gg F1 \gg M2 \gg F2$
VII.	(876, 504, 1288, 1334, 1286, 874, 1332)	$F4 \gg M3 \gg F3 \gg M1 \gg F1 \gg M2 \gg F2$
VIII.	(873, 0, 1255, 1000, 2000, 872, 2000)	$M1 \gg M3 \gg F3 \gg F4 \gg F1 \gg M2 \gg F2$
IX.	(869, 217, 1353, 1000, 2000, 867, 1349)	$M1 \gg F3 \gg M3 \gg F4 \gg F1 \gg M2 \gg F2$
X.	(922, 0, 1428, 1000, 1426, 920, 2000)	$M3 \gg F3 \gg M1 \gg F4 \gg F1 \gg M2 \gg F2$
XI.	(790, 285, 1227, 1146, 2001, 788, 1145)	$M1 \gg F3 \gg F4 \gg M3 \gg F1 \gg M2 \gg F2$
XII.	(1089, 1, 1336, 1000, 1088, 922, 2001)	$M3 \gg F3 \gg F1 \gg M1 \gg F4 \gg M2 \gg F2$
XIII.	(521, 452, 1339, 1292, 1291, 1337, 1218)	$F3 \gg M2 \gg F4 \gg M1 \gg M3 \gg F1 \gg F2$
XIV.	(439, 441, 1326, 1328, 1322, 1323, 1323)	$F4 \gg F3 \gg M2 \gg M3 \gg M1 \gg F2 \gg F1$
XV.	(290, 334, 1290, 1193, 2001, 1286, 1192)	$M1 \gg F3 \gg M2 \gg F4 \gg M3 \gg F2 \gg F1$
XVI.	(301, 316, 1200, 1217, 2001, 1199, 1193)	$M1 \gg F4 \gg F3 \gg M2 \gg M3 \gg F2 \gg F1$
XVII.	(334, 0, 1334, 1000, 2000, 1332, 2000)	$M1 \gg M3 \gg F3 \gg M2 \gg F4 \gg F1 \gg F2$
XVIII.	(357, 215, 1357, 1000, 2001, 1283, 1355)	$M1 \gg F3 \gg M3 \gg M2 \gg F4 \gg F1 \gg F2$
XIX.	(445, 1, 1445, 1000, 1444, 1332, 2001)	$M3 \gg F3 \gg M1 \gg M2 \gg F4 \gg F1 \gg F2$

The final ranking returned by the model for the case of language XI is incorrect, because it lets [za] surface faithfully. Let me diagnose the problem encountered by the model with the language XI. In order to get started, let me characterize the rankings that OT-correspond to this language XI using the two conditions in (236). The entire phonotactic comparative tableau corresponding to the language XI is given in (287).

	F_1	F_2	F_3	F_4	M_1	M_2	M_3
(pa, pa, ba)	W		W				W
(ba, ba, pa)	W		W				L
(ap, ap, ab)	W						W
(ab, ab, ap)	W						L
(sa, sa, za)		W		W			W
(as, as, az)		W					W
(apsa, apsa, absa)	W				W	W	
(apsa, apsa, apza)		W		W	W		W
(apsa, apsa, abza)	W	W		W		W	W
(abza, abza, apza)	W				W	L	
(abza, abza, absa)		W		W	W		L
(abza, abza, apsa)	W	W		W		L	L
(aspa, aspa, asba)	W		W		W	W	
(aspa, aspa, azpa)		W			W		W
(aspa, aspa, azba)	W	W	W			W	W
(azba, azba, azpa)	W		W		W	L	
(azba, azba, asba)		W			W		L
(azba, azba, aspa)	W	W	W			L	L

In order to check condition (236a), we can simplify the tableau (287) as follows. First, we can eliminate all the rows that do not contain a single w. Furthermore, we can eliminate: the row

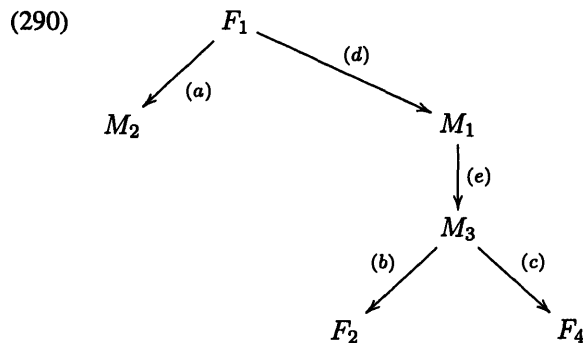
corresponding to the triplet (ba, ba, pa), because it is entailed by the row corresponding to the triplet (ab, ab, ap); the row corresponding to the triplet (abza, abza, absa), because it is entailed by the row corresponding to the triplet (azba, azba, absa); the row corresponding to the triplet (azba, azba, azpa), because it is entailed by the row corresponding to the triplet (abza, abza, apza); the row corresponding to the triplet (abza, abza, apza), because it is entailed by the row corresponding to the triplet (ab, ab, ap); the entries equal to L corresponding to M_2 in the two rows corresponding to the triplets (abza, abza, apsa) and (azba, azba, aspa), because they are entailed by the row corresponding to the triplet (ab, ab, ap). In conclusion, a ranking is OT-compatible with the comparative tableau (287) iff it is OT-compatible with the comparative tableau (288).

$$(288) \quad \left[\begin{array}{cccc|ccc} F_1 & F_2 & F_3 & F_4 & M_1 & M_2 & M_3 \\ \hline W & & & & & L & \\ W & W & & W & & & L \\ & W & & & W & & L \\ W & W & W & & & & L \end{array} \right]$$

Furthermore, I provide in (289) the list of blocks $A_{(\cdot, \cdot, z)}$ in the comparative tableau (287) corresponding to a loser form z that does not belong to the target language XI.

	F_1	F_2	F_3	F_4	M_1	M_2	M_3
(sa, sa, za)		W		W			W
(as, as, az)		W					W
(apsa, apsa, absa)	W				W	W	
(abza, abza, absa)		W		W	W		L
(apza, apza, absa)	W	W		W		W	L
(abza, abza, apza)	W				W	L	
(apsa, apsa, apza)		W		W	W		W
(absa, absa, apza)	W	W		W		L	W
(aspa, aspa, asba)	W		W		W	W	
(azba, azba, asba)		W			W		L
(azpa, azpa, asba)	W	W	W			W	L
(azba, azba, azpa)	W		W		W	L	
(aspa, aspa, azpa)		W			W		W
(asba, asba, azpa)	W	W	W			L	W

In order for condition (236a) to hold for the first row of the tableau (288), the ranking (290a) must hold. In order for condition (236b) to hold for the block $A_{(\cdot, \cdot, za)}$ in the tableau (289), the rankings (290b) and 290c) must hold. Thus, condition (236b) holds for the block $A_{(\cdot, \cdot, az)}$. In order for condition (236a) to hold for the second row of the tableau (288), the ranking (290d) must hold. The latter ranking ensures that condition (236a) holds for the last row of the tableau (288). In order for condition (236a) to hold for the third row of the tableau (288), the ranking (290e) must hold. The rankings already established ensure that condition (236b) holds for all remaining blocks $A_{(\cdot, \cdot, za)}$ in the tableau (289). In conclusion, a ranking OT-corresponds to the target language XI iff it satisfies the ranking conditions in (290).



The reason why the model fails on the language XI can now be spelled out as follows. According to (290), in order for a ranking to OT-correspond to the target language XI, it needs to rank the *general* faithfulness constraint F_1 above the unrelated *positional* faithfulness constraint F_4 . But this ranking configuration is hard for the model to achieve. In fact, the initial ranking vector ranks all positional faithfulness constraints above all general faithfulness constraints. As stated in the conjecture³ (291), if the initial separation θ^{init} between positional and general faithfulness constraints is large enough, then it will not be overcome during learning, namely the final ranking vector will still rank all positional faithfulness constraints above all general faithfulness constraints.⁴ Given a pair of a positional and a general faithfulness constraints that target the *same* feature (say IDENT[STOP-VOICING] and IDENT[STOP-VOICING]/ONSET), then it does make sense to only consider rankings where the positional faithfulness constraint is ranked above the corresponding general faithfulness constraint. But if the positional and the general faithfulness constraints target two different features (say IDENT[STOP-VOICING] and IDENT[FRICATIVE-VOICING]/ONSET), then there might be cases, like the case of language XI in (284), where we do want the general faithfulness constraint ranked above the *unrelated* positional faithfulness constraint. There are various conceivable ways to circumvent this problem with the initial bias of positional faithfulness constraints above general ones, that I plan to explore in future work.

- (291) For every typology, there exists an initial ranking vector that ranks positional faithfulness constraints above general faithfulness constraints with a separation large enough that that initial ranking will be maintained through learning for every target language in the typology.

Apart from the problem with language XI just discussed, I think it is remarkable that the model does get right all remaining 18 cases. What exactly ensures this success? In the rest of this subsection, I address this question. Let me start by making the question more explicit. In section 4.3, I have reviewed from the literature the idea of modeling the acquisition of phonology by means of the simple OT on-line algorithm (205). In particular, I have concentrated on the issue of modeling a special stage of the acquisition of phonology, namely the early stage described by Hayes (2004). Somewhat idealizing, this early stage is characterized by two crucial properties, summarized in (292). The *first* property is that the learner has no knowledge of the morphology of the target language \mathcal{L} throughout this entire stage. Since lack of morphological awareness means inability to detect alternations, then we can plausibly model this first property of the early stage by assuming that the underlying forms posited in step 1 of the OT on-line model are fully faithful. The *second* property is that the learner has acquired knowledge of the phonotactics of the target language \mathcal{L} by the end of this stage. We can model this second property by requiring that the final ranking \gg^{fin} entertained by the OT on-line model at the end of this stage OT-corresponds to the target language, namely satisfies the identity $\mathcal{R}(\text{OT}_{\gg^{\text{fin}}}) = \mathcal{L}$, that says that the language generated by the corresponding OT grammar $\text{OT}_{\gg^{\text{fin}}}$ coincides with the target language \mathcal{L} .

(292)	<i>properties of the early stage</i>	\implies	<i>and their modelization</i>
a.	lack of knowledge of the morphology of the target language \mathcal{L}	\implies	the underlying forms posited in step 1 of the OT on line model are fully faithful
b.	knowledge of the phonotactics of the target language \mathcal{L}	\implies	the final ranking \gg^{fin} returned by the OT on-line model satisfies the property $\mathcal{R}(\text{OT}_{\gg^{\text{fin}}}) = \mathcal{L}$

In subsection 4.3.4, I have noted that there is no way of modeling the early stage of the acquisition of phonology if we implement the OT on-line algorithm with a demotion-only update rule. In fact, if the algorithm posits fully faithful underlying forms, then the faithfulness constraints will never be loser-preferer. Thus, if the algorithm uses a demotion-only update rule, then the initial ranking of

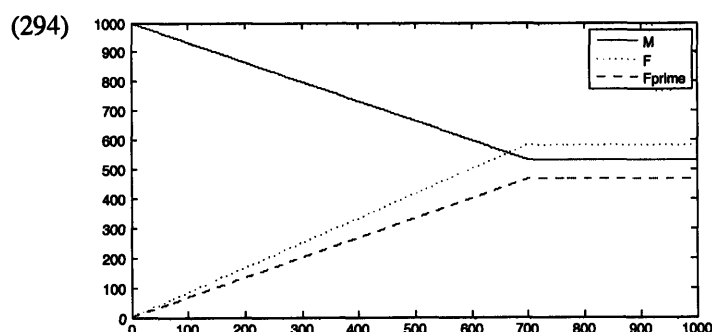
³I think that the conjecture (291) can easily be proven, but I defer the proof to future work.

⁴Indeed, both positional faithfulness constraints F_3 and F_4 are ranked above both general faithfulness constraints F_1 and F_2 in all the final ranking vectors reported in (286), with the only exception of the language XII. The latter exception is due to the fact that the initial separation $\theta^{\text{init}} = 1000$ between positional and general faithfulness constraints is too small.

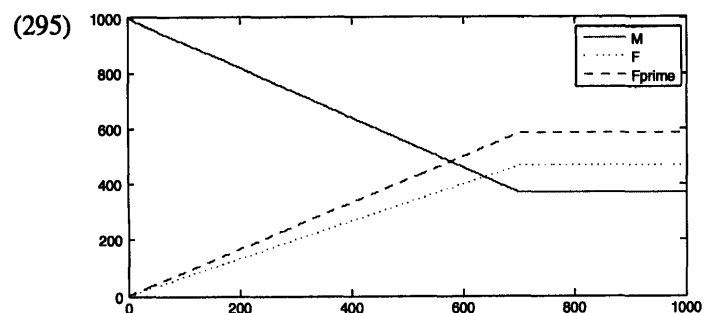
the faithfulness constraints will never be modified throughout learning. Hence, if the faithfulness constraints (of the same type) start out equally ranked, then the algorithm will never entertain the ranking configuration in (293), whereby a markedness constraint M is ranked in between two faithfulness constraints F, F' . Yet, a ranking configuration such as (293) might very well be needed in order to capture the phonotactics of the target language, namely in order for the final ranking \gg^{fin} to satisfy the condition $\mathcal{R}(\text{OT}_{\gg^{\text{fin}}}) = \mathcal{L}$.

$$(293) \quad F \gg M \gg F'$$

To overcome this impasse, in chapter 5 I have set myself the goal of devising new update rules for the OT on-line algorithm (205), that perform both promotion and demotion. Because of the promotion component of these new update rules, they are able to move around the faithfulness constraints even though they are never loser-preferrer. This means that the ranking configuration in (293) is *in principle* achievable starting with F and F' equally ranked, through the dynamics of the ranking values schematized in (294).



But will the ranking configuration (293) *indeed* be achieved? Let me make the issue explicit. It is indeed plausible that through learning F and F' will be pulled apart starting from their initial identical ranking value, because they will plausibly be promoted at different rates. If they get separated in such a way that F is above F' , then we expect that M will drop only as low as required to cross F , stopping before it crosses F' too, thus ending up in the target configuration (293). But why should F grow higher than F' ? The dynamics in (295) looks just as plausible: F' is promoted at a faster rate than F and is thus always above F ; in order for M to drop below F , it thus also has to drop below F' . In this case, even though the algorithm has pulled the two faithfulness constraints apart starting from their initial equal ranking value, the algorithm still ends up with the same unrestrictive configuration $F, F' \gg M$ that it would have obtained with a demotion-only update rule.



Thus, in order for the OT on-line algorithm with a promotion-demotion update rule to have a chance of working as a model of the early stage of the acquisition of phonology, something like the conjecture (296) must hold. In the rest of this section, I restate the conjecture (296) more explicitly, and discuss in detail its validity in the case of the Azba typology.

(296) Whenever the ranking configuration (293) is “needed” in order to account for the phonotactics of the target language, then it is indeed achieved by the model, namely constraint F is indeed promoted by the model at a faster rate than constraint F' .

Let me say that a given ranking \gg is a POSITIONAL-ABOVE-GENERAL (henceforth: PAG) ranking iff it ranks all positional faithfulness constraints above all general faithfulness constraints. To simplify the discussion, I will restrict myself to target languages that OT-correspond to PAG ranking. As noted above, in the case of the Azba typology, this restriction holds for all languages in the typology but for the problematic language XI. This restriction to languages that OT-correspond to PAG rankings reduces the number of cases that needs to be considered in discussing the conjecture (296). In fact, since the target language OT-corresponds to a PAG ranking, then no ranking configurations (293) with F a general faithfulness constraint and F' a positional faithfulness constraint can be needed. Furthermore, the reverse case where F is a positional faithfulness constraint and F' is a general faithfulness constraint will always be enforced. In fact, as noted in (291), the model will plausibly return a ranking vector with all positional faithfulness constraints ranked above all general faithfulness constraints, provided that the initial separation between the two sets of constraints is large enough. Thus, in assessing the conjecture (296) for target languages that OT-correspond to a PAG ranking, I only have to worry about the case where both F and F' are positional faithfulness constraints or else they are both general faithfulness constraints. Thus, let me restate the conjecture (296) as in (297).

(297) Whenever the ranking configuration (293) is “needed” (with F, F' both positional or both general) in order to account for the phonotactics of a target language OT-corresponding to a PAG ranking, then this configuration is indeed achieved by the model, namely constraint F is indeed promoted by the model at a faster rate than constraint F' .

The condition $M \gg F'$ in (293) can mean one of two things. Either M is ranked immediately before F' ; or else M is ranked above some other constraint C and this other constraint C is ranked immediately above F' , as in (298). In this intermediate constraint C is a faithfulness constraint F'' , then we can forget about F' and consider the ranking configuration (293) with F' replaced by F'' , namely $F \gg M \gg_{\text{im}} F''$. If this intermediate constraint C is a markedness constraint M' , then we can forget about M and consider the the ranking configuration (293) with M replaced by M' , namely $F \gg M' \gg_{\text{im}} F''$. In either case, we can always assume that the intermediate markedness constraint M in the ranking configuration (293) is ranked *immediately* above the lower faithfulness constraint F' .

(298) $F \gg \dots M \gg \dots C \gg_{\text{im}} F'$

I can thus restate the conjecture (297) more explicitly as in (299). Note that the notion of “being crucial” adopted here is extremely weak. In fact, it is a notion of being crucial relative to a specific ranking. An immediate ranking might be crucial for a given ranking OT-corresponding to the target language but not crucial for some other ranking OT-corresponding to the target language. The fact that I am using a rather weak notion of “being crucial” makes of course the conjecture (299) rather stronger.

(299) Consider a target language \mathcal{L} that OT-corresponds to a PAG ranking \gg . Assume that there are two faithfulness constraints F, F' and a markedness constraint M such that the two immediate rankings $F \gg_{\text{im}} \dots M \gg_{\text{im}} F'$ are crucial in \gg w.r.t. \mathcal{L} as defined in (233). Then this configuration is indeed achieved by the model, namely constraint F is indeed promoted by the model at a faster rate than constraint F' .

Does conjecture (299) hold in the general case? In the rest of this section, I start to tackle this question, by showing that it holds for the case of the Azba typology. Here, I concentrate on the case where F, F' are the two positional faithfulness constraints F_3, F_4 . The case where F, F' are the two general faithfulness constraints F_1, F_2 is slightly more complicated, and I defer to future work a through discussion of this case. By exploiting the symmetry between the two features STOP-VOICING and FRICATIVE-VOICING, it is sufficient to consider the case $F_3 \gg \dots \gg F_4$, since the case $F_3 \gg \dots \gg F_4$ then holds too by symmetry. Thus, conjecture (299) in this case becomes (300).

(300) Consider a target language \mathcal{L} in the Azba typology that OT-corresponds to a PAG ranking \gg . Assume that the two immediate rankings $F_3 \gg_{\text{im}} \dots M_i \gg_{\text{im}} F_4$ are crucial in \gg w.r.t.

\mathcal{L} for one of the three markedness constraints M_i with $i = 1, 2, 3$. Then this configuration is indeed achieved by the model, namely constraint F_3 is indeed promoted by the model at a faster rate than constraint F_4 .

The relative speed with which the two positional faithfulness constraints F_3 and F_4 raise over time depends on two factors: the number of rows in the input phonotactic tableau that promote one of the two but not the other; and the frequency with which those rows are sampled. In this section, I am adopting the oversimplification that the rows of the input phonotactic comparative tableau are sampled *uniformly*. This is of course an implausible assumption. But it provides a good starting point for the analysis of the model, because it makes the relative promotion speed of the two faithfulness constraints F and F' only depend on their number of W's in the input tableau. Let me say that a form y is an F_3 -PUSHER iff the corresponding block $A_{(y,y,\cdot)}$ has the following property: it contains a row that has a W in correspondence of F_3 but no W in correspondence of F_4 , but not vice versa. And of course F_4 -PUSHERS are defined analogously. As shown by the comparative tableau corresponding to the maximal language in (301), F_3 -pushers are [pa], [ba], [azpa], [aspa], [asba], [azba]; and F_4 -pushers: pushers are [sa], [za], [apza], [apsa], [absa], [abza].

	F_1	F_2	F_3	F_4	M_1	M_2	M_3
(pa, pa, ba)	W		W			W	
(ba, ba, pa)	W		W			L	
(ap, ap, ab)	W					W	
(ab, ab, ap)	W					L	
(sa, sa, za)		W		W			W
(za, za, sa)		W		W			L
(as, as, az)		W					W
(az, az, as)		W					L
(apza, apza, abza)	W				L	W	
(apza, apza, apsa)		W		W	L		L
(apza, apza, absa)	W	W		W		W	L
(apsa, apsa, absa)	W				W	W	
(apsa, apsa, apza)		W		W	W		W
(apsa, apsa, abza)	W	W		W		W	W
(abza, abza, apza)	W				W	L	
(abza, abza, absa)		W		W	W		L
(abza, abza, apsa)	W	W		W		L	L
(absa, absa, apsa)	W				L	L	
(absa, absa, abza)		W		W	L		W
(absa, absa, apza)	W	W		W		L	W
(azpa, azpa, azba)	W		W		L	W	
(azpa, azpa, aspa)		W			L		L
(azpa, azpa, asba)	W	W	W			W	L
(aspa, aspa, asba)	W		W		W	W	
(aspa, aspa, azpa)		W			W		W
(aspa, aspa, azba)	W	W	W			W	W
(azba, azba, azpa)	W		W		W	L	
(azba, azba, asba)		W			W		L
(azba, azba, aspa)	W	W	W			L	L
(asba, asba, aspa)	W		W		L	L	
(asba, asba, azba)		W			L		W
(asba, asba, azpa)	W	W	W			L	W

Thus, under the assumption that the rows of the input tableau are sampled uniformly, the conjecture (300) really becomes the following claim 32.

Claim 32 Consider a target language \mathcal{L} in the Azba typology that OT-corresponds to a PAG ranking \gg . Assume that the two immediate rankings $F_3 \gg_{\text{im}} \dots M_i \gg_{\text{im}} F_4$ are crucial in \gg w.r.t. \mathcal{L} for

one of the three markedness constraints M_i with $i = 1, 2, 3$. Then, the target language contains more F_3 -pushers than it contains F_4 -pushers.

Proof of the case M_1 . Assume that the target language OT-corresponds to a PAG ranking \gg such that the two immediate rankings $F_3 \gg_{\text{im}} \dots \gg_{\text{im}} M_1 \gg_{\text{im}} F_4$ are crucial. Let me show that the target language then has the properties (302), that show that it contains more F_3 -pushers than F_4 -pushers.

- (302) a. F_3 -pushers:
 the target language must contain [ba] and [azba];
 the target language cannot contain neither [azpa] nor [asba].
- b. F_4 -pushers:
 the target language can contain [za]
 the target language cannot contain [apza], [abza], or [absa].

By claim 30, in order for the immediate ranking $M_1 \gg_{\text{im}} F_4$ to be crucial, the phonotactic comparative tableau corresponding to the target language must contain a row (x, x, y) such that $y \notin \mathcal{L}$, both M_1 and F_4 have a w in that row and $\text{dec}_{\gg}(x, x, y) = M_1$. As shown in (301), there are only two possible rows where both M_1 and F_4 have a w, namely the rows corresponding to (abza, abza, absa) and to (apsa, apsa, apza). In either case, the condition $\text{dec}_{\gg}(x, x, y) = M_1$ means that the ranking \gg satisfies the conditions (303). Thus, the target language cannot contain [absa], because \gg could not be OT-compatible with the row (absa, absa, abza); cannot contain [apza], because \gg could not be OT-compatible with the row (apza, apza, apsa); cannot contain [azpa], because \gg could not be OT-compatible with the row (azpa, azpa, aspa); cannot contain [asba], because \gg could not be OT-compatible with the row (asba, asba, azba).

- (303) $M_1 \gg \{F_2, F_4, M_3\}$.

Again by claim 30, the constraint crucially ranked immediately below F_3 must be a markedness constraint (no immediate ranking of two faithfulness constraints can be crucial). It cannot be the constraint M_3 , since we have just concluded that M_3 is ranked below M_1 and M_1 is ranked below F_3 . Thus, the constraint crucially ranked immediately below F_3 can be either M_1 or M_2 . In other words, the ranking \gg must contain either the crucial immediate rankings (304a) or those in (304b).

- (304) a. $F_3 \gg_{\text{im}} M_2 \gg M_1 \gg_{\text{im}} F_4$.
 b. $F_3 \gg_{\text{im}} M_1 \gg_{\text{im}} F_4$

If case (304a) holds, then the target language contains [azba], because \gg is OT-compatible with the block $\mathbf{A}_{(\text{azba}, \text{azba}, \cdot)}$; and it contains [ba], because \gg is OT-compatible with the row (ba, ba, pa); but it cannot contain [abza], because \gg cannot be OT-compatible with the row corresponding to (abza, abza, apsa), since M_2 is \gg -ranked above F_4 and thus also above F_1 and F_2 by the hypothesis that \gg is a PAG. If case (304b) holds, then the ranking $F_3 \gg M_1$ is crucial. Claim 30 thus entails that the comparative tableau corresponding to the target language contains a row where F_3 has a w and M_1 has an L. The only two such rows are those corresponding to the triplets (azpa, azpa, azba) and (asba, asba, aspa). But the input comparative tableau cannot contain neither of these two rows, because we just concluded that the target language cannot contain neither [azpa] nor [asba]. ■

Proof of the case M_2 . Assume that the target language OT-corresponds to a PAG ranking \gg such that the two immediate rankings $F_3 \gg_{\text{im}} \dots \gg_{\text{im}} M_2 \gg_{\text{im}} F_4$ are crucial. Let me show that the target language then has the properties (305), that show that it contains more F_3 -pushers than F_4 -pushers.

- (305) a. F_3 -pushers:
 the target language contains [ba]
 if the target language contains [za], then it contains [azba].
- b. F_4 -pushers:
 the target language cannot contain [apza], [abza], or [absa]
 the target language may contain [za].

Let me start by proving claim (305b). By claim 30, in order for the immediate ranking $M_2 \gg_{\text{im}} F_4$ to be crucial, there has got to exist a form y such that the three conditions (306) hold.

- (306) a. $y \notin \mathcal{L}$
 b. there is $x \in \mathcal{L}$ such that $\text{Dec}_{\gg}(x, x, y) = M_2$ and F_4 has a w in (x, x, y) ;
 c. for any other $z \in \text{Gen}(y) \cap \mathcal{L}$, we have $\text{Dec}_{\gg}(z, z, y) \in \mathcal{F}$.

As shown in (301), there are only two possible rows where both M_2 and F_4 have a w and thus condition (306b) holds, namely the rows corresponding to the triplets (apsa, apsa, abza) and (apza, apza, absa). In either case, the condition $\text{dec}_{\gg}(x, x, y) = M_2$ entails that the ranking \gg satisfies the condition (307).

$$(307) \quad M_2 > \{F_1, F_2, F_4, M_3\}.$$

Thus, the target language cannot contain [abza], because the ranking \gg would not be OT-compatible with the row (abza, abza, apsa). Furthermore, the target language cannot contain [absa], because the ranking \gg would not be OT-compatible with the row (absa, absa, apza). Finally, the target language cannot contain [apza]. In fact, assume by contradiction that the target language did contain [apza]. Then, the ranking \gg would have to be OT-compatible with the row (apza, apza, abza) and thus it should satisfy the condition $M_1 \ll \max\{M_2, F_1\}$. Since $M_2 \gg F_1$ by (307), then $\max\{M_2, F_1\} = M_2$ and thus the latter condition becomes (308). But then condition (306c) fails both for the case of $y = [\text{abza}]$, because $\text{dec}_{\gg}(\text{apza}, \text{apza}, \text{abza}) = M_2$; and also for the case $y = [\text{absa}]$, because $\text{dec}_{\gg}(\text{apsa}, \text{apsa}, \text{absa}) = M_2$.

$$(308) \quad M_1 \ll M_2$$

Let me now turn to the proof of (305a). Assume that the target language contains [za], and thus that $F_4 \gg M_3$. Since the target language does not contain neither [abza] nor [absa], then the form [apza] needs to be neutralized to [apsa]. As shown by the row (apsa, apsa, apza), this means that one of the two rankings in (309) must hold. The ranking (309a) cannot hold, because of the assumption that the target language contains [za] and thus $F_4 > M_3$. Thus, the ranking in (309b) must hold.

- (309) a. $M_3 \gg \{F_2, F_4, M_1\}$
 b. $M_1 \gg \{F_2, F_4, M_3\}$

The two rankings (307) and (309b) together with $F_3 \gg M_2$ entail that the ranking \gg is OT-compatible with all three rows in the block $\mathbf{A}_{(\text{azba}, \text{azba}, \cdot)}$ and thus the target language must contain [azba]. ■

Proof of the case M_3 . Assume that the target language OT-corresponds to a PAG ranking \gg such that the two immediate rankings $F_3 \gg_{\text{im}} \dots \gg_{\text{im}} M_3 \gg_{\text{im}} F_4$ are crucial. Let me show that the target language then has the properties (310), that show that it contains more F_3 -pushers than F_4 -pushers.

- (310) a. F_3 -pushers:
 if the target language contains [absa], then it also contains [asba] and [ba]
 b. F_4 -pushers:
 the target language cannot contain [za], [apza], or [abza];
 the target language might contain [absa].

Since $M_3 \gg F_4$, then the target language cannot contain [za], since the ranking \gg could not be OT-compatible with the row (za, za, sa); nor can it contain [apza], since the ranking \gg could not be OT-compatible with the row (apza, apza, apsa). Since $M_3 \gg F_4$ and furthermore \gg is a PAG, then the target language cannot contain [abza], because of the row (abza, abza, apsa). Assume that the target language contains [absa]. Then, the ranking \gg must satisfy the two conditions (311).

- (311) a. $F_1 \gg M_1, M_2$
 b. $M_1 \ll \max\{F_2, F_4, M_3\} = M_3$

These two conditions (311) ensure that the ranking \gg is OT-compatible with the block $\mathbf{A}_{(\text{asba}, \text{asba}, \cdot)}$ and with the row (ba, ba, pa), so that the target language must contain both [asba] and [ba]. ■

Chapter 7

How to study constraint promotion: number of updates

In chapter 5, I have presented various update rules for the OT on-line algorithm (205) that can trigger only a finite number of updates, despite the fact that they perform a certain amount of promotion too. In chapter 6, I have started the investigation of the properties of these promotion-demotion update rules. In particular, I have addressed the issue (206b) concerning the characterization of the final vector returned by the OT on-line algorithm run with such update rules. I have presented a dual invariant for the algorithm, and I have used it to study its performance on “naturalistic” input phonotactics comparative tableaux. I have put forward the conjecture that the sensitivity of promotion-demotion update rules to the *w*’s in these tableaux does indeed help the algorithm to converge to the correct final ranking. In this chapter, I pursue further the investigation of promotion-demotion update rules, by addressing the issue (206c) concerning the number of updates required by the OT on-line algorithm run with such update rules.

- (206) c. Given a specific update rule, what is the worst case number of updates performed in step 3 by the OT on-line algorithm (205), over all possible input comparative tableaux with n columns and all possible ways of sampling the rows of the input comparative tableau?

As reviewed in subsection 4.2.3, minimal demotion-only update rules can only perform a number of updates small w.r.t. the number n of constraints. In section 7.1, I show that update rules that perform promotion too might on the contrary require a large number of updates. There is thus a sharp contrast (312) between demotion-only and promotion-demotion update rules. The somewhat obvious diagnosis of the contrast in (312) is as follows. A promotion-demotion update rule is sensitive to the *w*’s in the input comparative tableau, while a demotion-only update rule is not. An input comparative tableau might contain many *w*’s that have nothing to contribute to the final OT-compatible ranking vector. One way to construct such tableaux is to take a diagonal tableau as in (56) and to add *w*’s at the right of its *L*’s. These extra *w*’s are of course completely superfluous, namely the tableau thus obtained is OT-equivalent to the initial diagonal tableau, namely both of them are only OT-compatible with the unique ranking $C_1 \gg C_2 \gg \dots \gg C_n$. Section 7.1 shows that in such cases, the sensitivity of promotion-demotion update rules to the *w*’s in the input tableaux ends up being just a distraction that slows down the algorithm. Demotion-only update rules are not sensitive to the *w*’s in the input tableau, and thus cannot be slowed down by useless distractions.

- (312) a. *Demotion-only update rules:*
the worst case number of updates is always small (namely quadratic in n).
b. *Promotion-demotion update rules:*
the worst case number of updates can be large (namely exponential in n).

How can we make sense of the slowness (312b) of promotion-demotion update rules? how does their slowness (312b) bear on the issue of their evaluation? Here is a preview of the answer I want to

suggest in the rest of this chapter. The task of the acquisition of phonology can of course be stated as a formal, explicit computational problem. This statement will have something like the shape (313).

- (313) *given*: a comparative tableau $\mathbf{A} \in \bigcup_{m,n} \{\text{L, E, W}\}^{m \times n}$ that satisfies certain properties $\mathcal{P}_{\text{given}}$;
find: a ranking \gg that satisfies certain properties $\mathcal{P}_{\text{find}}$.

The problem statement (313) depends on two properties. The property $\mathcal{P}_{\text{given}}$ carves out of the set $\bigcup_{m,n} \{\text{L, E, W}\}^{m \times n}$ of all possible comparative tableaux the special subset of those that are relevant to the statement of the problem. This property $\mathcal{P}_{\text{given}}$ might correspond to something like the loose statement in (314). The property $\mathcal{P}_{\text{find}}$ establishes the criterion of success in order for a ranking \gg (or, equivalently, a ranking vector θ) to represent a solution of the problem.

- (314) A comparative tableau \mathbf{A} satisfies property $\mathcal{P}_{\text{given}}$ iff \mathbf{A} corresponds to some *realistic* set of data corresponding to the *realistic* universal specifications that underly the typology of natural language phonology.

A model for the acquisition of phonology is of course a function that takes an arbitrary input comparative tableau $\mathbf{A} \in \bigcup_{m,n} \{\text{L, E, W}\}^{m \times n}$ and returns a ranking \gg (or a ranking vector θ). One measure of theoretical significance of a given model is its worst case running time. The worst case can be computed over two different sets of inputs, as stated in (315): over the entire set of comparative tableaux; or over the subset of those comparative tableaux that satisfy $\mathcal{P}_{\text{given}}$. The measure (315a) is of obvious theoretical significance. The measure (315b) might turn out to be an interesting alternative, if we are not able to provide an explicit description of the property $\mathcal{P}_{\text{given}}$, along the lines of (314).

- (315) a. What is the worst case running time of the algorithm over the set of comparative tableaux that satisfies property $\mathcal{P}_{\text{given}}$?
 b. What is the worst case running time of the algorithm over the entire set of comparative tableaux, no matter whether they satisfy property $\mathcal{P}_{\text{given}}$ or not?

The OT on-line algorithm (205) discussed so far is of course an instance of such a model. And its running time can be identified with its number of updates. The result presented in section 7.1 and previewed in (312) looks at the issue from the point of view of (315b): it says that the worst case number of updates over the set of all comparative tableaux is exponential in the number of columns. In this chapter, I want to argue that this fact should not be held against promotion-demotion update rules. In section 7.2, I will review from the literature various formulations of the computational problem (313) of the acquisition of phonology. And in section 7.4 I will make the point that, at least when the property $\mathcal{P}_{\text{given}}$ in (314) is dropped from the formulation of the problem, then *every* algorithm for these problems has worst case running time (315b) exponential in the number of columns of the given comparative tableau. What about the perspective in (315a)? As noted above, the idea behind claim (312b) is that constraint promotion can be fooled by constructing input tableaux with many superfluous W's that slow the algorithm down since they just constitute a distraction. Yet, the test cases discussed in the chapter 6, and especially in subsection 6.3.3, suggest the conjecture that, once we restrict ourselves to realistic cases that satisfy the property $\mathcal{P}_{\text{given}}$ in (314), the sensitivity of promotion-demotion update rules to the W's of the input tableaux does not in any way constitute a "distraction" but rather leads the algorithm to the right solution.

7.1 Constraint promotion might require many updates

In subsection 4.2.3.1, I reviewed the answer to question (206c) provided in the literature for the case of minimal demotion-only update rules. The core idea was to slightly refine the line of reasoning used in the proof of claim 4 to answer question (206a) in order to get an answer to question (206c) too, as shown in the proof of claims 5 and 6. This strategy turned out to be remarkably successful, since the worst case bound thus obtained was tight and only quadratic in the number n of constraints, as recalled in (312a). Also the two strategies presented in sections 5.2 and 5.4 to answer question

(206a) for promotion-demotion update rules can be refined to obtain actual bounds on the worst case number of updates. Subsection 7.1.3 illustrates this point for the case of the analysis presented in section 5.4. Yet, neither of these two analyses provides a bound on the worst case number of updates polynomial in the number n of constraints. It turns out that this failure is not due to a weakness of the analyses. In fact, the two subsections 7.1.1 and 7.1.2 show that there are indeed input comparative tableaux such that even the corresponding *best* case number of updates (over all possible ways of feeding rows in step 1) is exponential in the number n of constraints, as stated in (312b). The analysis is independent on the details of the specific promotion-demotion update rule considered.

7.1.1 A simple lower-bound on the best-case number of updates

To simplify the discussion, I will assume once more that the input comparative tableau \mathbf{A} contains exactly one entry equal to L per row. Claim 2 guarantees that this auxiliary assumption does not affect the generality of the analysis, since a general comparative tableau can be preprocessed and turned into an OT-equivalent comparative tableau with a unique entry equal to L per row. I will concentrate on the promotion-demotion update rule introduced in section 5.1, that in this case takes the form (123), repeated in (316).

(316) If the current ranking vector θ^{old} is not OT-compatible with a comparative row \mathbf{a} that contains a unique entry equal to L :

$$\text{a. } \theta_k^{\text{new}} \doteq \begin{cases} \theta_k^{\text{old}} + 1 & \text{if } k \in W(\mathbf{a}) \\ \theta_k^{\text{old}} - w(\mathbf{a}) & \text{if } k = L(\mathbf{a}) \\ \theta_k^{\text{old}} & \text{otherwise} \end{cases}$$

b. Promote all winner-preferring constraints by 1; demote the unique loser-preferring constraint by the total number $w(\mathbf{a})$ of winner-preferring constraints.

Let $\bar{\mathbf{a}}$ be the update vector corresponding to the update rule (316) w.r.t. update triggered by the row \mathbf{a} of the input comparative tableau, as defined in (161) and repeated in (317): each entry equal to W in \mathbf{a} gets replaced by 1 in $\bar{\mathbf{a}}$; each entry equal to E gets replaced by 0; each entry equal to L gets replaced by $-w(\mathbf{a})$, where $w(\mathbf{a})$ is the number of entries equal to W in the comparative row \mathbf{a} . Let $\bar{\mathbf{A}}$ be the numerical matrix obtained by organising these numerical vectors $\bar{\mathbf{a}}$ one underneath the other. Let $\bar{\mathbf{A}}_1, \dots, \bar{\mathbf{A}}_n \in \mathbb{R}^m$ be the n columns of the matrix $\bar{\mathbf{A}}$.

$$(317) \quad \mathbf{a} = [a_1 \quad \dots \quad a_k \quad \dots \quad a_n] \implies \bar{\mathbf{a}} = [\bar{a}_1 \quad \dots \quad \bar{a}_k \quad \dots \quad \bar{a}_n]$$

$$\bar{a}_k = \begin{cases} +1 & \text{if } a_k = W \\ 0 & \text{if } a_k = E \\ -w(\mathbf{a}) & \text{if } a_k = L \end{cases}$$

The following claim 33 provides a lower-bound on the best-case number of updates. The proof is a straightforward generalization of the reasoning used in section 4.2.6 to explain Pater's counterexample. For the sake of simplicity, claim 33 only considers the case of input comparative tableaux OT-compatible with a unique ranking; the case of compatible tableaux OT-compatible with multiple rankings is handled analogously: consider one such ranking at the time, compute the solution of the corresponding problem (318) and let the lower-bound on the best case number of updates be the smallest such solution. In the next subsection 7.1.2, I will use the bound provided by claim 33 to show that the best-case number of updates in the case of the promotion-demotion update rule (316) can be exponential in n .

Claim 33 Consider a comparative tableau $\mathbf{A} \in \{L, E, W\}^{m \times n}$ that contains a unique entry equal to L per row and that is OT-compatible with a unique ranking. Without loss of generality, assume that this ranking is $C_1 \gg C_2 \gg \dots \gg C_n$ (otherwise, relabel the constraints). The number of updates performed by the on-line algorithm (205) with the promotion-demotion update rule (316) run on this input tableau \mathbf{A} starting from the null initial vector cannot be smaller than the solution of the program (318) in the decision variable $\alpha \in \mathbb{R}^m$.

$$(318) \quad \begin{aligned} &\text{minimize: } \sum_{i=1}^m \alpha_i \\ &\text{subject to: } \langle \bar{\mathbf{A}}_k - \bar{\mathbf{A}}_{k+1}, \boldsymbol{\alpha} \rangle \geq 1 \quad k = 1, \dots, n-1 \\ &\quad \boldsymbol{\alpha} \geq \mathbf{0} \end{aligned}$$

Proof. Consider a generic run of the algorithm on the input comparative tableau \mathbf{A} . As proved in sections 5.2 or 5.4, the algorithm will converge to a ranking vector OT-compatible with the input comparative tableau \mathbf{A} after a finite number of updates. Let me introduce the following pieces of notation: let T be the total number of updates required by the algorithm to converge; let $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ be the final ranking vector OT-compatible with \mathbf{A} returned by the algorithm; for every row $i = 1, \dots, m$ of the input comparative tableau \mathbf{A} , let $\alpha_i \in \mathbb{N}$ be the final dual variable associated with the i th row, namely the number of times in that run that the algorithm has updated its current ranking vector in response to a failure in accounting for the i th row of that tableau; let $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)$ be the corresponding final dual vector. Let me recall two important relationships between these three characters T , $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$. The identity (319) trivially holds: the total number of updates T is the sum of the number of updates α_1 triggered by first row, plus the number of updates α_2 triggered by the second row, etcetera.

$$(319) \quad T = \sum_{i=1}^m \alpha_i$$

Furthermore, the identity (320) holds for every $k = 1, \dots, n$: the k th component θ_k of the final ranking vector is the scalar product between the k th column $\bar{\mathbf{A}}_k$ of the derived numerical matrix $\bar{\mathbf{A}}$ and the final dual vector $\boldsymbol{\alpha}$. This identity (320) is a straightforward generalization of the identity (95) discussed in detail in section 4.2.6 in the context of the explanation of Pater's counterexample.

$$(320) \quad \theta_k = \langle \bar{\mathbf{A}}_k, \boldsymbol{\alpha} \rangle$$

Since the input comparative tableau is only OT-compatible with the ranking $C_1 \gg C_2 \gg \dots \gg C_n$, then the final vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ must satisfy the $n - 1$ strict inequalities (321a) in order to be OT-compatible with the input comparative tableau. By (320), these primal inequalities (321a) can be rewritten in terms of the dual variables as in (321b). By the linearity of the scalar product $\langle \cdot, \cdot \rangle$, these inequalities (321b) can equivalently be rewritten as in (321c). Finally, since the dual vector $\boldsymbol{\alpha}$ has integer components (because each component represents the number of times that update has been triggered by the corresponding row) and since the entries of the derived numerical matrix $\bar{\mathbf{A}}$ in (317) are all integer, then these strict inequalities (321c) are equivalent to the loose inequalities (321d).

$$(321) \quad \begin{aligned} &\text{For } k = 1, \dots, n-1: \\ &\text{a. } \theta_k > \theta_{k+1} \\ &\text{b. } \langle \bar{\mathbf{A}}_k, \boldsymbol{\alpha} \rangle > \langle \bar{\mathbf{A}}_{k+1}, \boldsymbol{\alpha} \rangle \\ &\text{c. } \langle \bar{\mathbf{A}}_k - \bar{\mathbf{A}}_{k+1}, \boldsymbol{\alpha} \rangle > 0 \\ &\text{d. } \langle \bar{\mathbf{A}}_k - \bar{\mathbf{A}}_{k+1}, \boldsymbol{\alpha} \rangle \geq 1 \end{aligned}$$

By virtue of (319) and (321), the total number of updates performed by the algorithm is lower bounded by the solution of the integer program (322).

$$(322) \quad \begin{aligned} &\text{minimize: } \sum_{i=1}^m \alpha_i \\ &\text{subject to: } \langle \bar{\mathbf{A}}_k - \bar{\mathbf{A}}_{k+1}, \boldsymbol{\alpha} \rangle \geq 1 \quad k = 1, \dots, n-1 \\ &\quad \boldsymbol{\alpha} \in \mathbb{N}^m \end{aligned}$$

The claim follows by relaxing the integer constraint $\boldsymbol{\alpha} \in \mathbb{N}^m$ of the integer program (322) into the linear constraint $\boldsymbol{\alpha} \geq \mathbf{0}$ of the linear program (318). \blacksquare

7.1.2 A case with an exponential best-case number of updates

Claim 33 says that we can use the solution to the linear program (318) as a lower-bound on the best-case number of updates of the OT on-line algorithm (205) with the promotion-demotion update rule (316). In other words, it says that if there is a family of comparative tableaux $\{A_n\}_{n \in \mathbb{N}}$ with n columns for every $n = 1, 2, \dots$ such that the solution of the linear program (318) corresponding to the tableau A_n grows exponentially in n , then the number of updates required by the on-line algorithm (116) with the promotion-demotion update rule (316) will grow exponentially with n . In this subsection, I construct one such family of comparative tableaux $\{A_n\}_{n \in \mathbb{N}}$. For every n , let $A_n \in \{L, E, W\}^{(n-1) \times n}$ be the comparative tableau with $n - 1$ rows and n columns obtained from the diagonal tableau in (56) by “adding” two extra entries equal to W at the right of every diagonal entry equal to L . To illustrate, I give A_7 in (323).

$$(323) \quad A_7 = \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 \\ \begin{matrix} W \\ \\ \\ \\ \\ \\ \end{matrix} & L & W & W & & & & \\ & W & L & W & W & & & \\ & & W & L & W & W & & \\ & & & W & L & W & W & \\ & & & & W & L & W & \\ & & & & & W & L & \end{matrix}$$

The MatLab code `MinimumRunningTime` (available on the author’s website) takes a comparative tableau $A \in \{L, E, W\}^{m \times n}$ compatible with the unique ranking $C_1 \gg C_2 \gg \dots \gg C_n$, constructs the corresponding linear program (318) and solves it using the built-in Matlab LP-solver. I give in (324) the solutions of the linear program (318) corresponding to comparative tableaux of the form (323) for various values of n , together with a corresponding solution.¹

(324)	$n = 5$	$n = 7$	$n = 9$	$n = 11$
optimal value of the LP (318)	56	594	5,036	38,413
and corresponding solution α	$\begin{bmatrix} 2 \\ 7 \\ 19 \\ 28 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 11 \\ 31 \\ 79 \\ 194 \\ 276 \end{bmatrix}$	$\begin{bmatrix} 4 \\ 15 \\ 43 \\ 111 \\ 275 \\ 670 \\ 1,622 \\ 2,296 \end{bmatrix}$	$\begin{bmatrix} 5 \\ 19 \\ 55 \\ 143 \\ 356 \\ 870 \\ 2,110 \\ 5,102 \\ 12,323 \\ 17,430 \end{bmatrix}$

The data in (324) show that the best-case number of updates required by the on-line algorithm (116) with the promotion-demotion update rule (316) can be super-polynomial in n .

7.1.3 A small remark on Pater’s comparative tableaux

In the preceding subsection, I have used the counterexample of the comparative tableaux illustrated in (323) to show that the best-case number of updates required by the on-line algorithm (116) with the promotion-demotion update rule (316) can be super-polynomial in n . These comparative tableaux are obtained by adding *two* W ’s immediately at the right of the L of each row of the diagonal comparative tableau (56). The intuition is that these two extra W ’s do not in any way contribute to the OT-compatibility of the tableau and thus they just end up “distracting” the algorithm, which therefore takes much longer to converge. In this subsection, I note that adding only *one* distracting W per row is not able to catastrophically affect the behavior of the algorithm. The goal of the discussion is to show how the proof of finite time convergence presented in section 5.4 can be trivially refined

¹It is interesting to note that the solution on the linear program is integral for every n , so that the relaxation done at the end of the proof of claim 33 is not needed. I do not know why that is.

to obtain a bound on the actual number of updates. Let PATER'S TABLEAU of order n be the comparative tableau $\mathbf{A}_n \in \{L, E, W\}^{n-1 \times n}$ with $n - 1$ rows and n columns obtained from the diagonal tableau in (56) by "adding" a *single* entry equal to w at the right of every diagonal entry equal to L . The case with $n = 5$ was given in (82), and is repeated in (325). For convenience, in this subsection I label the constraints starting from the right-most column of the comparative tableau.

$$(325) \quad \begin{matrix} & C_5 & C_4 & C_3 & C_2 & C_1 \\ \begin{bmatrix} W & L & W & & \\ & W & L & W & \\ & & W & L & W \\ & & & W & L \end{bmatrix} \end{matrix}$$

Claim 34 shows that the algorithm requires a number of update polynomial in n in the case of Pater's comparative tableaux. For completeness, I provide in (326) the lower bound provided by the solution of linear program (318) corresponding to Pater's comparative tableau for various values of n . These data show that the bound provided by claim 34 is rather loose.

(326)	$n = 5$	$n = 7$	$n = 9$	$n = 11$
optimal value of the LP (318)	25	98	270	605
and corresponding solution α	$\begin{bmatrix} 2 \\ 5 \\ 8 \\ 10 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 8 \\ 14 \\ 20 \\ 25 \\ 28 \end{bmatrix}$	$\begin{bmatrix} 4 \\ 11 \\ 20 \\ 30 \\ 40 \\ 49 \\ 56 \\ 60 \end{bmatrix}$	$\begin{bmatrix} 5 \\ 14 \\ 26 \\ 40 \\ 55 \\ 70 \\ 84 \\ 96 \\ 105 \\ 110 \end{bmatrix}$

Claim 34 The on-line algorithm (205) with the promotion-demotion update rule (316) run on Pater's comparative tableau of order n starting from the null initial vector can perform a maximum number of updates of the order of n^5 .

Proof. Consider the derived numerical matrix $\bar{\mathbf{A}}$ corresponding to the input Pater's tableau \mathbf{A} as in (317). The line of reasoning presented in section 5.4, based on claim 25 reviewed in the Appendix, immediately yields the bound (327) on the number of updates of the on-line algorithm (205) run on that input tableau \mathbf{A} with the promotion-demotion update rule (316).

$$(327) \quad \# \text{ of updates} \leq \left(\frac{\text{maximum 2-norm of the rows of } \bar{\mathbf{A}}}{\text{best-margin of the rows of } \bar{\mathbf{A}}} \right)^2$$

The maximum 2-norm of the rows of $\bar{\mathbf{A}}$ can straightforwardly be bound by n^2 . The best-margin of the rows of $\bar{\mathbf{A}}$ is the quantity defined in (328). Thus, in order to use the bound (327), we need a lower bound the corresponding best margin (328). In the rest of this proof, I explicitly compute (328) for the case where the input comparative tableau \mathbf{A} is Pater's tableau of order n .

$$(328) \quad \text{best-margin of the rows of } \bar{\mathbf{A}} = \max_{\theta \neq 0} \min_{\bar{\mathbf{a}}} \frac{\langle \theta, \bar{\mathbf{a}} \rangle}{\|\theta\|}$$

It is well known that the best margin (328) can equivalently be described as the square root of the inverse of the solution of the quadratic optimization problem (329); see for instance Vapnik (1998, Theorem 10.2).

$$(329) \quad \begin{aligned} &\text{minimize: } \|\theta\|^2 \\ &\text{subject to: } \langle \theta, \bar{\mathbf{a}} \rangle \geq 1 \quad \text{for every row } \bar{\mathbf{a}} \end{aligned}$$

A row $\bar{\mathbf{a}}$ of the derived matrix $\bar{\mathbf{A}}$ is called a SUPPORT VECTOR iff the constraint $\langle \theta, \bar{\mathbf{a}} \rangle \geq 1$ in the definition of the feasible set in (329) holds tight at optimality, namely $\langle \theta^*, \bar{\mathbf{a}} \rangle = 1$, where θ^*

is the unique solution of the optimization problem (329). Note that in the special case of Pater's comparative tableau, all rows \bar{a} are support vectors, so that the optimization problem (329) can be rewritten as in (330). In fact, suppose by contradiction that there exists a row of the derived matrix \bar{A} that is not a support vector; for concreteness, suppose that it is the first row \bar{a}_1 (the reasoning immediately extends to an arbitrary row). This means that for any vector $\theta \in \mathbb{R}^n$, if $\langle \theta, \bar{a} \rangle \geq 1$ for every row $\bar{a} \neq \bar{a}_1$, then also $\langle \theta, \bar{a}_1 \rangle \geq 1$. Consider an arbitrary nonnegative vector $\theta = (\theta_1, \dots, \theta_n)$ with $\theta_1 = 0$ such that $\langle \theta, \bar{a} \rangle \geq 1$ for every row \bar{a} , which exists by claim 20. Consider next the vector $\theta' = (\theta'_1, \dots, \theta'_n)$ which is identical to θ but for the fact that $\theta'_n = 0$. Of course, $\langle \theta', \bar{a} \rangle \geq 1$ for every row \bar{a} but the first one (because $\langle \theta', \bar{a} \rangle = \langle \theta, \bar{a} \rangle$ for every such row, since every row \bar{a} but the first one has null first component). If it were also $\langle \theta', \bar{a}_1 \rangle \geq 1$, then θ' would be OT-compatible with Pater's tableau, by claim 19. But this cannot be, because the ranking $C_2 \gg \dots \gg C_n \gg C_1$ is a refinement of θ' and it is obviously not OT-compatible with Pater's tableau.

$$(330) \quad \begin{array}{ll} \text{minimize:} & \|\theta\|^2 \\ \text{subject to:} & \langle \theta, \bar{a} \rangle = 1 \quad \text{for every row } \bar{a} \end{array}$$

It turns out that the optimization problem (330) is very easy to solve. Consider for instance the case of Pater's comparative tableau of order $n = 5$, repeated in (331) together with the corresponding numerical matrix \bar{A} . Note that there exists a unique ranking vector $\theta = (\theta_1, \dots, \theta_5)$ such that $\theta_1 = 0$ and furthermore $\langle \theta, \bar{a} \rangle = 1$ for every row \bar{a} of the corresponding numerical matrix \bar{A} , namely the vector in (332). Just as in the proof of claim 20, this vector is obtained by starting from the bottom row of \bar{A} and moving upward using the condition $\langle \theta, \bar{a}_k \rangle = 1$ for the k th row from the bottom to univocally determine θ_{k+1} .

$$(331) \quad \begin{array}{l} \text{row 4} \\ \text{row 3} \\ \text{row 2} \\ \text{row 1} \end{array} \begin{array}{ccccc} C_5 & C_4 & C_3 & C_2 & C_1 \\ \left[\begin{array}{ccccc} W & L & W & & \\ & W & L & W & \\ & & W & L & W \\ & & & W & L \end{array} \right] \end{array} \implies \bar{A} = \begin{bmatrix} +1 & -2 & +1 & & \\ & +1 & -2 & +1 & \\ & & +1 & -2 & +1 \\ & & & +1 & -1 \end{bmatrix}$$

$$(332) \quad \begin{array}{l} \text{row 1} \\ \text{row 2} \\ \text{row 3} \\ \text{row 4} \end{array} \begin{array}{l} \implies \theta_2 - \theta_1 = 1 \\ \implies \theta_3 - 2\theta_2 + \theta_1 = 1 \\ \implies \theta_4 - 2\theta_3 + \theta_2 = 1 \\ \implies \theta_5 - 2\theta_4 + \theta_3 = 1 \end{array} \begin{array}{l} \theta_1 = 0 \\ \implies \theta_2 = 1 \\ \implies \theta_3 = 3 \\ \implies \theta_4 = 6 \\ \implies \theta_5 = 10 \end{array}$$

In the general case, there is a unique ranking vector θ OT-compatible with Pater's comparative tableau of order n such that $\theta_1 = 0$ and $\langle \theta, \bar{a} \rangle = 1$ for every row \bar{a} of the derived numerical matrix, and it is defined by the recursion in (333).

$$(333) \quad \begin{array}{l} \theta_1 = 0 \\ \theta_2 = 1 \\ \theta_k = 1 + 2\theta_{k-1} - \theta_{k-2} \end{array}$$

It is trivial to prove by induction on k that the recursion (333) can be made explicit as in (334) for every $k = 1, \dots, n$.

$$(334) \quad \theta_k = \frac{k^2 - k}{2}$$

So far, I have made only one arbitrary choice, namely setting θ_1 to zero. I can fix this arbitrariness as follows. Let e be the vector of \mathbb{R}^n with all components equal to 1. Note that all the rows of \bar{a} are orthogonal to e , namely $\langle e, \bar{a} \rangle = 0$. Thus, given a nonnegative vector $\theta \geq 0$ such that $\langle \theta, \bar{a} \rangle = 1$ for every row \bar{a} , we also have $\langle \theta + \xi e, \bar{a} \rangle = 1$, for every row \bar{a} and for any choice of the constant $\xi \in \mathbb{R}$. We can thus pick ξ such that the 2-norm of the corresponding vector $\theta + \xi e$ is minimized. In other words, we have to find ξ that minimizes the function $f(\xi) = \sum_{k=1}^n (\theta_k + \xi)^2$. Since the function $f(\xi)$ is strictly convex, the problem is solved by setting to zero the derivative of f , which yields (335).

$$(335) \quad \xi^* = -\frac{1}{n} \sum_{k=1}^n \theta_k = -\frac{1}{n} \|\theta\|_1$$

The (squared) 2-norm of the corresponding vector $\theta + \xi e$ is computed in (336)

$$(336) \quad \begin{aligned} \|\theta + \xi^* e\|^2 &= \sum_{k=1}^n \left(\theta_k - \frac{1}{n} \|\theta\|_1 \right)^2 \\ &= \sum_{k=1}^n \left(\theta_k^2 - \frac{1}{n^2} \|\theta\|_1^2 - \frac{2}{n} \theta_k \|\theta\|_1 \right) \\ &= \|\theta\|^2 - \frac{1}{n} \|\theta\|_1^2 \end{aligned}$$

In conclusion, the best margin of the rows of the matrix \bar{A} derived from Pater's comparative tableau of order n can be computed as in (337), by using (336) and (334) respectively.

$$(337) \quad \begin{aligned} \left(\frac{1}{\text{best-margin of the rows of } \bar{A}} \right)^2 &= \|\theta\|^2 - \frac{1}{n} \|\theta\|_1^2 \\ &= \sum_{k=2}^n \left(\frac{k^2 - k}{2} \right)^2 - \frac{1}{n} \sum_{k=2}^n \frac{k^2 - k}{2} \end{aligned}$$

The leading term of the final expression obtained in (337) is $\sum_{k=1}^n k^4$ and thus the claim follows from the well-known identity $\sum_{k=1}^n k^4 = \frac{1}{5}n^5 + \frac{1}{2}n^4 + \frac{1}{3}n^3 - \frac{1}{30}n$. ■

7.2 Formulations of the problem of the acquisition of phonology

In this section, I review from the literature three different formulations of the problem (313) of the acquisition of phonology within OT.

7.2.1 The Ranking problem

Consider again the example of the universal specifications in (1), repeated in (338). A language in the corresponding typology is given in (339), corresponding to the ranking $F_{\text{pos}} \gg M \gg F_{\text{gen}}$.

$$(338) \quad \begin{aligned} \text{a. } \mathcal{X} &= \{ /ta/, /da/, /rat/, /rad/ \} \\ \text{b. } \mathcal{Y} &= \{ [ta], [da], [rat], [rad] \} \\ \text{c. } \text{Gen}(/ta/) &= \text{Gen}(/da/) = \{ [ta], [da] \} \\ \text{Gen}(/rat/) &= \text{Gen}(/rad/) = \{ [rat], [rad] \} \\ \text{d. } \mathcal{C} &= \left\{ \begin{array}{l} F_{\text{pos}} = \text{IDENT}[\text{ONSET}][\text{VOICE}], \\ F_{\text{gen}} = \text{IDENT}[\text{VOICE}], \\ M = *[\text{+VOICE}, \text{-SONORANT}] \end{array} \right\} \end{aligned}$$

$$(339) \quad \{ [da], [ta], [rat] \}$$

Consider a learner that is exposed to the language (339). Upon hearing many instances of the surface forms [da] and [rat], he might decide that these forms do indeed belong to the target language he is being exposed to. As noted in subsection 4.3.2, under mild conditions on the constraint set, it makes sense to assume that these forms are the faithful surface realization of the corresponding identical underlying forms. Thus, the learner might assume that the ranking \gg that corresponds to the target language he is being exposed to validates the pairs of an underlying form and a corresponding winner surface form in (340a). Or perhaps the learner will have had access to some alternations and will thus have noticed that the grammar he is being exposed to maps the underlying form /rad/ to the surface form [rat], thus neutralizing final voicing. In this case, the learner will posit the set of underlying/winner form pairs in (340b). For the discussion in the rest of this chapter, it is not crucial how exactly the learner will get to a set of data such as those in (340), or which one of these

two sets of data the learner will actually get to. The general point is just that the learner needs to have some device to construct suitable underlying/winner form pairs such as those in (340) from the language (339) he is being exposed to; see for instance Tesar and Smolensky (2000, Ch. 4) for much elaboration on this point.

- (340) a. $\mathcal{D} = \{(/da/, [da]), (/rat/, [rat])\}$
 b. $\mathcal{D} = \{(/da/, [da]), (/rad/, [rat])\}$

Suppose that the universal specifications \mathcal{X} , \mathcal{Y} , Gen and \mathcal{C} in (338) are fixed and known a priori. Then the learner needs at the very least to solve problem (341). The learner is provided with a set of data \mathcal{D} , namely a set of pairs of an underlying form together with the corresponding intended winner surface form. His task is to come up with a ranking of the constraint set that “accounts” for the data \mathcal{D} , namely such that the corresponding OT-grammar maps each underlying form in \mathcal{D} into the corresponding surface form.

- (341) *given:* a set of data \mathcal{D} , as in (340);
find: a ranking \gg of the constraint set \mathcal{C} that “accounts” for the set of data \mathcal{D} .

Let me generalize the specific problem (341) into a general statement. Consider a set \mathcal{D} of DATA, namely a finite set of pairs (x, y^*) of an underlying form $x \in \mathcal{X}$ and a corresponding intended winner candidate surface form $y^* \in Gen(x) \subseteq \mathcal{Y}$. For future reference, it is useful to have in place the straightforward definition (342) of OT-COMPATIBILITY between one such set of data and an arbitrary ranking.

- (342) a. A ranking \gg is called OT-COMPATIBLE with a data set \mathcal{D} iff the corresponding OT-grammar OT_{\gg} corresponding to that ranking \gg accounts for all pairs in \mathcal{D} , namely $OT_{\gg}(x) = y^*$ for every pair $(x, y^*) \in \mathcal{D}$.
 b. The set of data \mathcal{D} is called OT-COMPATIBLE iff \mathcal{D} is OT-compatible with at least a ranking according to (342a).

Suppose that some universal specifications \mathcal{X} , \mathcal{Y} , Gen and \mathcal{C} are fixed and known a priori. I can thus generalize the specific problem (341) into the general statement (343). This formulation (343) assumes the set of data \mathcal{D} to be OT-compatible. This is of course unrealistic, because the data set \mathcal{D} might very well contain a small number of corrupted pairs that make \mathcal{D} not OT-compatible. But I will ignore this issue here, and stick with the assumption that the input set of data is indeed OT-compatible.²

- (343) *given:* a finite OT-compatible set of data $\mathcal{D} \subseteq \mathcal{X} \times \mathcal{Y}$;
find: a ranking \gg of the constraint set \mathcal{C} that is OT-compatible with \mathcal{D} .

We would like to list problem (343) among the relevant problems for the acquisition of phonology. But of course, problem (343) would be useful to this end only if we did have knowledge of the actual universal specifications $(\mathcal{X}, \mathcal{Y}, Gen, \mathcal{C})$. At this stage of the development of the field, of course we don't. Thus, problem (343) as it stands is of no use. To overcome this difficulty, we change the problem statement (343) to (344), by letting the universal specifications figure as a variable parameter of the problem. Thus, we don't need to worry anymore about what the actual universal specifications look like, because we will require the model to work no matter what. This is of course not a small modification. This modification really makes sense only if we can confidently assume that a proper model of the acquisition of phonology should not rely on properties of the specific, actual universal specifications. As we will see in subsection 7.4.1, this assumption will turn out to be warranted and the switch from (343) to (344) harmless. This trick of coping with the lack of knowledge of the actual universal specifications by quantifying universally over universal specifications, is standard in the OT complexity literature. For example, Eisner (2000) writes: “we follow Tesar and Smolensky (2000) in supposing that the learner already knows the correct set of

²To relax this assumption, we would have to decide what to ask for in the case \mathcal{D} is not OT-compatible. If we only ask that OT-incompatibility be detected, then we get a variant of the problem (343) that is not much harder. If we ask instead for a ranking OT-compatible with the largest number of pairs in \mathcal{D} , then we get a much harder problem.

constraints \mathcal{C} . The assumption follows from the OT philosophy that \mathcal{C} is universal across languages, and only the [ranking] of constraints differ. The algorithms for learning a ranking, however, are designed to be general for any \mathcal{C} , so they take \mathcal{C} as an input. That is, these methods are not tailored (as others might be) to exploit the structure of some specific, putatively universal [constraint set] \mathcal{C} ." Problem (344) is called the RANKING PROBLEM.

- (344) *given:* a) universal specifications \mathcal{X} , \mathcal{Y} , Gen and \mathcal{C} ;
 b) a finite OT-compatible set of data $\mathcal{D} \subseteq \mathcal{X} \times \mathcal{Y}$;
find: a ranking \gg of the constraint set \mathcal{C} that is OT-compatible with \mathcal{D} .

7.2.2 The Subset problem

Can we stop here? Is (344), or at least (343), a fair characterization of the task of the acquisition of phonology within OT? No, it is not. Here is a way to appreciate the point. Suppose that the data set \mathcal{D} provided as input to the problem (344) is (340a). The corresponding Ranking problem (344) admits the two solutions (345). These two solutions generate two different languages. The language in (345b) is a proper subset of the language in (345a). A number of authors have suggested that the ranking that corresponds to the small language (345b) is a better solution of the given instance of the Ranking problem than the ranking (345a) that corresponds to the large grammar; see for instance Berwick (1985), Manzini and Wexler (1987), Prince and Tesar (2004) and Hayes (2004) for discussion.

- (345) a. $F_{pos} \gg F_{gen} \gg M \implies \mathcal{R}(OT_{\gg}) = \{ta, da, rat, rad\}$
 b. $F_{pos} \gg M \gg F_{gen} \implies \mathcal{R}(OT_{\gg}) = \{ta, da, rad\}$

Let me encode this intuition by turning to the more demanding variant of the Ranking problem (344) stated in (346). Suppose that some universal specifications \mathcal{X} , \mathcal{Y} , Gen and \mathcal{C} are fixed and known a priori. Again, the learner is given a data set \mathcal{D} , that for simplicity is assumed to be OT-compatible. But now the learner's task is not just to come up with *any* ranking OT-compatible with \mathcal{D} . Rather, the learner needs to come up with one such ranking that furthermore has the property that the corresponding language is as small as possible, among the languages generated by rankings OT-compatible with \mathcal{D} .

- (346) *input:* a finite OT-compatible data set $\mathcal{D} \subseteq \mathcal{X} \times \mathcal{Y}$;
output: a ranking \gg OT-compatible with \mathcal{D} such that there exists no other ranking \gg' OT-compatible with \mathcal{D} too such that $\mathcal{R}(OT_{\gg'}) \subsetneq \mathcal{R}(OT_{\gg})$

As already noted above in the case of problem (343), also problem (346) is of little use as a test case for models of the acquisition of phonology, as long as we cannot make explicit the actual universal specifications. To overcome this difficulty, we modify problem (346) as in (347), by letting the universal specifications figure once more as an arbitrary input to the problem. Problem (347) is called the SUBSET PROBLEM.

- (347) *input:* a) universal specifications \mathcal{X} , \mathcal{Y} , Gen and \mathcal{C} ;
 b) a finite OT-compatible data set $\mathcal{D} \subseteq \mathcal{X} \times \mathcal{Y}$;
output: a ranking \gg OT-compatible with \mathcal{D} such that there exists no other ranking \gg' OT-compatible with \mathcal{D} too such that $\mathcal{R}(OT_{\gg'}) \subsetneq \mathcal{R}(OT_{\gg})$

Consider the special case where the learner is provided with an entire language \mathcal{L} in the typology corresponding to some universal specifications $(\mathcal{X}, \mathcal{Y}, Gen, \mathcal{C})$. Assume furthermore that $\mathcal{X} = \mathcal{Y}$ and that the learner assumes fully faithful corresponding underlying forms. In other words, the learner is given a set of data \mathcal{D} that coincides with the diagonal of $\mathcal{L} \times \mathcal{L}$. In this case, the requirement on the output ranking vector in (347) boils down to the condition $\mathcal{R}(OT_{\gg}) = \mathcal{L}$ that the ranking \gg OT-corresponds to the language \mathcal{L} . Thus, the problem considered in section 6.2 is a special case of the Subset problem (347).

7.2.3 Prince and Tesar's (2004) formulation of the Subset problem

Prince and Tesar (2004) offer an interesting alternative formulation of the Subset problem (347). The idea behind their reformulation can be introduced as follows. Let a STRICTNESS MEASURE be a function μ that takes a ranking \gg and returns a number $\mu(\gg) \in \mathbb{N}$ that provides a relative measure of the size of the language $\mathcal{R}(\text{OT}_{\gg})$ corresponding to \gg , in the sense that the (strict) monotonicity property in (348) holds for any two rankings \gg, \gg' : if the language $\mathcal{R}(\text{OT}_{\gg'})$ corresponding to \gg' is a proper subset of the language $\mathcal{R}(\text{OT}_{\gg})$ corresponding to \gg , then the strictness measure of \gg' is strictly smaller than the strictness measure of \gg . Thus, the smaller the strictness measure, the smaller the language.

$$(348) \quad \mathcal{R}(\text{OT}_{\gg'}) \subset \mathcal{R}(\text{OT}_{\gg}) \implies \mu(\gg') < \mu(\gg).$$

Let μ be a strictness measure. Any solution of problem (349) is guaranteed to be a solution of the Subset problem (347). In fact, if a ranking \gg is a solution of problem (349), then there cannot exist any other ranking \gg' OT-compatible with \mathcal{D} that corresponds to a language $\mathcal{R}(\text{OT}_{\gg'})$ such that $\mathcal{R}(\text{OT}_{\gg'}) \subsetneq \mathcal{R}(\text{OT}_{\gg})$, since (348) would then imply that $\mu(\gg') < \mu(\gg)$, thus contradicting the hypothesis that \gg is a solution of problem (349).

$$(349) \quad \begin{array}{l} \text{input:} \quad \text{a) universal specifications } \mathcal{X}, \mathcal{Y}, \text{Gen and } \mathcal{C}; \\ \quad \quad \quad \text{b) a finite set } \mathcal{D} \subseteq \mathcal{X} \times \mathcal{Y}; \\ \text{output:} \quad \text{a ranking } \gg \text{ OT-compatible with } \mathcal{D} \text{ such that there exists no other ranking} \\ \quad \quad \quad \gg' \text{ OT-compatible with } \mathcal{D} \text{ too such that } \mu(\gg') < \mu(\gg) \end{array}$$

The reformulation (349) looks promising. Yet, we now need to come up with a strictness measure. Of course, not just any strictness measure will do. For instance, the function (350), which pairs a ranking \gg with the cardinality of the corresponding language $\mathcal{R}(\text{OT}_{\gg})$, trivially satisfies (348), and is thus a strictness measure. Yet, this is not a *good* strictness measure, because there seems to be no way to compute $\mu(\gg)$ without actually computing the corresponding language $\mathcal{R}(\text{OT}_{\gg})$.

$$(350) \quad \mu(\gg) = \text{cardinality of the language } \mathcal{R}(\text{OT}_{\gg}).$$

Prince and Tesar (2004) suggest a better candidate. As usual, assume that the set of constraints $\mathcal{C} = \mathcal{F} \cup \mathcal{M}$ is split up into the subset \mathcal{F} of faithfulness constraints and the subset \mathcal{M} of markedness constraints. They put forward the intuition (351). The rationale behind this intuition is that faithfulness constraints work toward preserving the contrasts present in the set \mathcal{X} of underlying forms and thus a large language is likely to arise from high ranked faithfulness constraints.

$$(351) \quad \text{A ranking } \gg \text{ corresponds to a smallest language } \mathcal{R}(\text{OT}_{\gg}) \text{ if it ranks the faithfulness constraints "as low as possible".}$$

Based on this intuition (351), Prince and Tesar suggest the definition in (352). For each faithfulness constraint $F \in \mathcal{F}$, determine the number $\mu(F)$ of markedness constraints $M \in \mathcal{M}$ that are \gg -ranked *below* that faithfulness constraint and add up all these numbers $\mu(F)$ together to determine the value $\mu(\gg)$ for the ranking \gg . I will call problem (349) with the specific measure μ defined in (352) the PTSUBSET PROBLEM.

$$(352) \quad \mu(\gg) \doteq \sum_{F \in \mathcal{F}} \underbrace{\left| \{M \in \mathcal{M} \mid F \gg M\} \right|}_{\mu(F)}$$

According to (352), the μ measure of the ranking in (345a) is 2 while that of the desired ranking in (345b) is 1. Unfortunately, the function μ defined in (352) is *not* a strictness measure in the general case. Thus, the PTSubset (349) is only an "approximate" reformulation to the Subset problem (347).

7.3 How to decide whether a problem is easy or not

Given two sets \mathcal{J} and \mathcal{G} , a corresponding PROBLEM is any relation $\Pi \subseteq \mathcal{J} \times \mathcal{G}$ between the two sets \mathcal{J} and \mathcal{G} . Any element in the set \mathcal{J} is called an INSTANCE of the problem; every element in the set

\mathcal{S} is called a SOLUTION of the problem. A problem $\Pi \subseteq \mathcal{I} \times \mathcal{S}$ can also be represented as in (353). The Ranking problem (344), the Subset problem (347), and the PTSubset problem (349) illustrate the general scheme (353).

- (353) *given:* an instance $x \in \mathcal{I}$;
output: a solution $y \in \mathcal{S}$ such that $\Pi(x, y)$.

Assume that the set \mathcal{I} of instances comes with a function $|\cdot| : \mathcal{I} \rightarrow \mathbb{N}$ that pairs each instance x of the problem Π with a number $|x|$ that expresses the SIZE of that instance and reflects its complexity. A problem $\Pi = (\Pi, |\cdot|)$ is called TRACTABLE iff it admits a polynomial-time SOLUTION ALGORITHM, namely an algorithm Solve_Π that satisfies condition (354). Intuitively, the “easy” problems are the tractable ones.

- (354) For every instance $x \in \mathcal{I}$, Solve_Π runs on input x in time polynomial in its size $|x|$ and returns a solution $y = \text{Solve}_\Pi(x)$ such that $\Pi(x, y)$.

In the next section of this chapter, I want to tackle question (355). In particular, I will show that the Subset problem (347) and the PTSubset problem (349) are *not* tractable, namely that any solution algorithm for these problems requires a large running time in the worst case.

- (355) Are the Ranking problem (344), the Subset problem (347) and the PTSubset problem (349) tractable or not?

It is relatively easy to show that a given problem is tractable: one just needs to exhibit a polynomial-time solution algorithm, in the sense of (354). On the contrary, it is not at all trivial to show that a given problem is *not* tractable. Naively, one would have to show that no polynomial-time solution algorithm exists. This strategy is of course not viable. An alternative, more sophisticated strategy has been devised in the computational literature. This alternative strategy is informally stated in (356). The rationale is as follows: if there exist non-tractable problems and if Π' is among the hardest problems, then Π' must be non-tractable; thus, if our problem Π is at least as hard as Π' , then our problem Π has got to be non-tractable too.

- (356) Suppose that non-tractable problems exist. To conclude that a given problem Π is non-tractable, show that that problem Π is at least as hard as some other problem Π' that is among the hardest problems.

In the rest of this section, I review how (356) is usually formalized, one step at the time; see Garey and Johnson (1979) and Cormen et al. (1990, Ch. 36) for details.

7.3.1 First step

A DECISION PROBLEM is a problem $\Pi \subseteq \mathcal{I} \times \mathcal{S}$ whose set of solutions is $\mathcal{S} = \{0, 1\}$. Intuitively, decision problems are those problems that only ask for a “yes” or a “no”. For instance, the following problem (357) is the decision problem corresponding to the original PTSubset problem (349). Non-tractability of the decision problem (357) entails non-tractability of the original PTSubset problem (349). In fact, if the original problem (349) were tractable, then the decision problem (357) would be tractable too, since I could solve the decision problem by finding a solution \gg of the original problem (349) and then checking whether $\mu(\gg)$ is smaller than k or not.

- (357) *given:* a) universal specifications $\mathcal{X}, \mathcal{Y}, \text{Gen}$ and \mathcal{C} ;
b) a finite OT-compatible set of data $\mathcal{D} \subseteq \mathcal{X} \times \mathcal{Y}$;
c) an integer k ;
output: “yes” iff there exists a ranking \gg OT-compatible with \mathcal{D} s.t. $\mu(\gg) \leq k$.

The preceding considerations immediately extend to the general case: a general problem Π can be paired up with a corresponding decision problem Π_{dec} in such a way that non-tractability of the decision problem Π_{dec} entails non-tractability of the original problem Π . Thus, let me restate the informal statement (356) as in (358).

- (358) Suppose that non-tractable decision problems exist. To conclude that a given problem Π is non-tractable, show that the corresponding decision problem Π_{dec} is at least as hard as some other decision problem Π' that is among the hardest decision problems.

Note that (358) is now fully stated in terms of decision problems. The next step toward a proper formalization of (358) consists of making explicit the condition that the decision problem Π_{dec} is at least as hard as some other decision problem Π' .

7.3.2 Second step

Given two decision problems $\Pi_1 \subseteq \mathcal{J}_1 \times \mathcal{G}_1$ and $\Pi_2 \subseteq \mathcal{J}_2 \times \mathcal{G}_2$, we say that Π_1 is NOT HARDER THAN or that it REDUCES TO Π_2 (in symbols: $\Pi_1 \leq_P \Pi_2$) iff there exists an algorithm $\text{Reduction}_{\Pi_1, \Pi_2}$ that satisfies condition (359). The algorithm $\text{Reduction}_{\Pi_1, \Pi_2}$ is called a REDUCTION of Π_1 to Π_2 . The two problems Π_1 and Π_2 are called EQUIVALENT iff both $\Pi_1 \leq_P \Pi_2$ and $\Pi_2 \leq_P \Pi_1$.

- (359) For every instance $x_1 \in \mathcal{J}_1$ of Π_1 , $\text{Reduction}_{\Pi_1, \Pi_2}$ runs on x_1 in time polynomial in its size $|x_1|$ and returns an instance $x_2 = \text{Reduction}_{\Pi_1, \Pi_2}(x_1) \in \mathcal{J}_2$ of Π_2 such that $\Pi_1(x_1) = 1 \iff \Pi_2(x_2) = 1$

The relation \leq_P is reflexive, anti-symmetric and transitive, namely it is a partial order among decision problems. If $\Pi_1 \leq_P \Pi_2$, then any polynomial-time solution algorithm Solve_{Π_2} for Π_2 can be turned into the trivial polynomial-time solution algorithm Solve_{Π_1} in (360) for Π_1 .

- (360) $\text{Solve}_{\Pi_1}(x_1)$
 1 compute $x_2 = \text{Reduction}_{\Pi_1, \Pi_2}(x_1)$
 2 return $\text{Solve}_{\Pi_2}(x_2)$

Thus, the issue of solving problem Π_1 has been reduced to the issue of solving problem Π_2 ; namely, problem Π_1 cannot be harder than problem Π_2 . In conclusion, in order to show that a given decision problem Π_2 is not tractable, it is sufficient to show that there is a decision problem Π_1 such that $\Pi_1 \leq_P \Pi_2$ and furthermore Π_1 is not tractable. I thus further formalize (358) as in (361).

- (361) Suppose that non-tractable decision problems exist. To conclude that a given problem Π is non-tractable, show that $\Pi' \leq_P \Pi_{\text{dec}}$, where Π_{dec} is the decision problem corresponding to Π and Π' is some decision problem that is among the hardest decision problems.

The next step toward a proper formalization of (361) consists of making explicit the assumption that there exist non-tractable decision problems.

7.3.3 Third step

The set of tractable decision problems is denoted by \mathcal{P} . More explicitly, a decision problem Π belongs to the class \mathcal{P} iff it admits an algorithm Solve_{Π} that satisfies condition (362). This condition (362) is a straightforward adaptation to the case of decision problems of the general condition (354).

- (362) For every instance $x \in \mathcal{J}$, Solve_{Π} runs on input x in time polynomial in its size $|x|$ and furthermore $\Pi(x) = 1$ iff $\text{Solve}_{\Pi}(x) = 1$

Another important class of decision problems is \mathcal{NP} : a decision problem Π belongs to the class \mathcal{NP} iff Π admits a polynomial-time VERIFICATION ALGORITHM, namely an algorithm Verify_{Π} that satisfies condition (363) for some polynomial p . For instance, the decision problem (357) belongs to the class \mathcal{NP} : given a ranking (encoded as a not too long boolean vector y), it is easy to decide whether its corresponding μ -measure (352) is smaller than k or not.

- (363) For every instance $x \in \mathcal{J}$, $\Pi(x) = 1$ iff there exists $y \in \{0, 1\}^{p(|x|)}$ such that Verify_{Π} runs on input (x, y) in time polynomial in the size $|x|$ and returns 1.

Given an arbitrary problem $\Pi \in \mathcal{NP}$, we can use the corresponding verification algorithm Verify_{Π} to construct the algorithm Solve_{Π} in (364). Of course, Solve_{Π} is a solution algorithm for Π with worst-case running time of the order of $2^{p(|x|)}$. Thus, \mathcal{NP} is the class of decision problems for which brute force search yields an exponential time solution algorithm.

```

(364) Solve $_{\Pi}(x)$ 
      1  answer  $\leftarrow 0$ 
      2  for every  $y \in \{0, 1\}^{p(|x|)}$ 
      3      if Verify $_{\Pi}(x, y) = 1$  then
      4          answer  $\leftarrow 1$ 
      5  return answer

```

Of course, $\mathcal{P} \subseteq \mathcal{NP}$, since any polynomial-time solution algorithm Solve $_{\Pi}$ for Π can be used as a verification algorithm. Do the two classes \mathcal{P} and \mathcal{NP} coincide?; namely: do all decision problems that admit an exponential time solution algorithm also admit a polynomial time solution algorithm? This question is currently open in the literature. Yet, there are many problems in \mathcal{NP} for which no polynomial-time solution algorithm is currently known and that we are thus tempted to assume not to belong to \mathcal{P} . The COMPLEXITY CONJECTURE (365) says that there are indeed problems in \mathcal{NP} that do not belong to \mathcal{P} , namely are not tractable, namely do not admit a polynomial-time solution algorithm.

(365) $\mathcal{P} \neq \mathcal{NP}$

The complexity conjecture (365) formalizes the crucial assumption in (361) that there exist non-tractable decision problems. The statement in (361) can thus be formalized as in (366).

(366) Suppose that $\mathcal{P} \neq \mathcal{NP}$. To conclude that a given problem Π is non-tractable, show that $\Pi' \leq_{\mathcal{P}} \Pi_{\text{dec}}$, where Π_{dec} is the decision problem corresponding to Π and Π' is some decision problem that is among the hardest decision problems.

The last step toward a proper formalization of (366) consists of making explicit the assumption that Π' is among the hardest possible decision problems.

7.3.4 Fourth step

A decision problem Π is called **HARD** iff the following condition (367) holds. This definition says that NP-hard problems are at least as hard as any problem in \mathcal{NP} . In other words, it says that, if we had a polynomial-time solution algorithm for even just one NP-hard problem, then we would have a polynomial-time solution algorithm for every problem in \mathcal{NP} . A decision problem is called **NP-COMPLETE** iff it is hard and furthermore belongs to the class \mathcal{NP} .

(367) $\Pi' \leq_{\mathcal{P}} \Pi$ for every decision problem $\Pi' \in \mathcal{NP}$.

In can thus conclude this section with the fully explicit restatement of (366) provided in (368).

(368) Suppose that $\mathcal{P} \neq \mathcal{NP}$. To conclude that a given problem Π is non-tractable, show that $\Pi' \leq_{\mathcal{P}} \Pi_{\text{dec}}$, where Π_{dec} is the decision problem corresponding to Π and Π' is some NP-complete decision problem.

In section 7.4, I will use (368) to show that both the PTSubset problem (349) and the Subset problem (347) are non-tractable.

7.4 The problem of the acquisition of phonology is “hard”

In section 7.2, I reviewed various explicit computational problems that provide at least a loose, simplified description of the task of the acquisition of phonology. The first problem considered was the Ranking problem (344). That problem is of course tractable, as reviewed in subsection 7.4.1. But in section 7.2, I could not stop at that problem, and had to refine it into a more demanding problem that imposes further requirements on the output ranking. I thus introduced the Subset problem (347) and its reformulation as the PTSubset problem (349). In subsections 7.4.2 and 7.4.3, I show that the latter two problems are not tractable.

7.4.1 Complexity of the Ranking problem

In this subsection, I quickly address the question of whether the Ranking problem (344) is “easy” or not. On the background of the preceding section, this question boils down to the question of whether the Ranking problem is tractable or not, namely whether it admits or not a polynomial-time solution algorithm in the sense of (354). In order to address this question, we need preliminarily to complete the statement (344) of the problem with the specification of the size of an instance of the problem, namely the parameters that we allow the running time of a solution algorithm to depend on. To this end, let the WIDTH of the generating function Gen on the data set \mathcal{D} given with an instance of the Ranking problem (344) be the number $width(\mathcal{D})$ defined in (369) as the cardinality of the largest candidate set over all underlying forms that appear in \mathcal{D} .

$$(369) \quad width(\mathcal{D}) = \max_{(x,y^*) \in \mathcal{D}} |Gen(x)|$$

As stated in (370), I assume that the size of a given instance of the Ranking problem (344) depends on three parameters: the cardinality $|\mathcal{C}|$ of the constraint set, the cardinality $|\mathcal{D}|$ of the data set, and the width of the generating function $width(\mathcal{D})$ on the given data set \mathcal{D} . It is uncontroversial that the size of a given instance of the Ranking problem should depend on the cardinalities $|\mathcal{C}|$ and $|\mathcal{D}|$. It is more delicate to let it depend on $width(\mathcal{D})$ too. The potential difficulty with this assumption is as follows: that $width(\mathcal{D})$ could be very large, potentially exponential in the number of constraints $|\mathcal{C}|$; thus, letting the size depend on $width(\mathcal{D})$ might make the problem too easy, by loosening up too much the tight dependence on $|\mathcal{C}|$. This difficulty would indeed arise in the case of the original formulation (343) of the Ranking problem, where we do not have control over the generating function Gen . But we have replaced that formulation (343) with the alternative formulation (344), where we have provided for our lack of knowledge of the relevant universal specifications by universally quantifying over universal specifications. Thus, the difficulty just discussed does not arise for this latter formulation, since an alleged solution algorithm is required to work also for cases where the number of constraints $|\mathcal{C}|$ is large but the width of the generating function $width(\mathcal{D})$ on the data set is small. For instance, all case studies considered in section 6.3 had this property.³

- (370) *given:* a) universal specifications $\mathcal{X}, \mathcal{Y}, Gen$ and \mathcal{C} ;
 b) a finite OT-compatible set of data $\mathcal{D} \subseteq \mathcal{X} \times \mathcal{Y}$;
find: a ranking \gg of the constraint set \mathcal{C} that is OT-compatible with \mathcal{D} ;
size: $\max\{|\mathcal{C}|, |\mathcal{D}|, width(\mathcal{D})\}$.

The set of data \mathcal{D} given with an instance (370) of the Ranking problem can of course be paired up with its corresponding comparative tableau, obtained by considering all underlying/winner/loser form triplets (x, y^*, y) corresponding to a pair (x, y^*) in the set \mathcal{D} and a loser candidate $y \in Gen(x)$ different from y^* . I can thus state the Ranking problem (370) in terms of comparative tableaux as in (371). The size of the Ranking problem (371) is given of course by the number n of columns and the number m of rows of the input comparative tableau \mathbf{A} . Consider the set of data \mathcal{D} given with an instance (370) of the classical formulation of the Ranking problem. The corresponding comparative tableau $\mathbf{A}_{\mathcal{D}}$ has n columns and a number m of rows that can of course be bound by $m \leq |\mathcal{D}|width(\mathcal{D})$, where $|\mathcal{D}|$ is the number of pairs in the data set \mathcal{D} and $width(\mathcal{D})$ is the width of the generating function on the set \mathcal{D} , as defined in (369). Thus, an instance of the classical Ranking problem (370) can be transformed through into a corresponding instance of the Ranking problem (371) with comparable size. In conclusion, in order to prove that the original problem (370) is tractable, it is sufficient to prove that the reformulation (371) in terms of comparative tableaux is tractable.

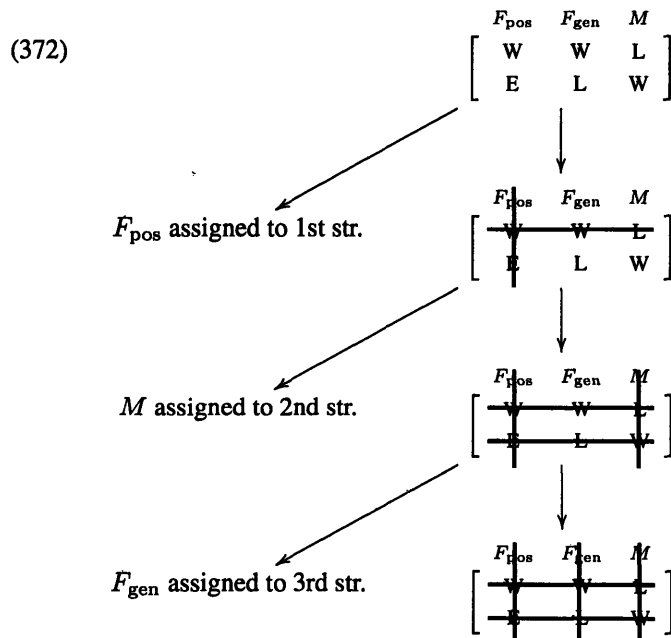
- (371) *given:* an OT-compatible comparative tableau $\mathbf{A} \in \{L, E, W\}^{m \times n}$;
find: a ranking \gg that is OT-compatible with the comparative tableau \mathbf{A} ;
size: $\max\{m, n\}$.

³Furthermore, letting the size of an instance of the Ranking problem depend on $width(\mathcal{D})$, as well as on the cardinalities $|\mathcal{C}|$ and $|\mathcal{D}|$, immediately ensures that the problem is in \mathcal{NP} , namely that it admits a polynomial time verification algorithm.

Tesar (1995), Tesar and Smolensky (1998) and Tesar and Smolensky (2000, Ch. 7) (henceforth: Tesar and Smolensky) prove the following important claim 35. This result has had a profound impact on the field.

Claim 35 The Ranking problem (371) is tractable.

A proof of claim 35 is provided by the quadratic running time of the OT on-line algorithm with demotion-only update rules. An alternative faster solution can be obtained as follows. Let me illustrate the idea as in (372) for the instance of the Ranking problem corresponding to the comparative tableau (15) repeated at the top of (372). Our goal is to come up with a ranking \gg -OT-compatible with the input tableau according to (13a), namely such that the tableau obtained by \gg -reordering the columns from left to right in decreasing order has the property that the leftmost non-E entry of each row is a W. The top ranked constraint must head a column that does not contain a single L. In our case, the only such constraint is F_{pos} , that thus gets assigned to the top stratum. The constraint that can be assigned to the next stratum must head a column whose only L's belong to rows where the top ranked constraint F_{pos} has a W. In other words, it must head a column that does not contain a single L once we strike out the rows where the top ranked constraint F_{pos} has a W. In our case, the only such constraint is M , that thus gets assigned to the second stratum. The constraint that can be assigned to the next stratum must head a column that does not contain a single L once we strike out rows where at least one of the two top ranked constraints F_{pos} and M have a W. In our case, the only such constraint is F_{gen} , that thus gets assigned to the bottom stratum.



The procedure just illustrated can be straightforwardly extended to the general case, thus obtaining the algorithm in (373). Step (373a) corresponds to the diagonal arrows in (372); step (373b) corresponds to the vertical arrows in (372). Algorithm (373) is due to Tesar and Smolensky and is called RECURSIVE CONSTRAINT DEMOTION (henceforth: RCD). Obviously, if the input comparative tableau A is OT-compatible, then RCD returns a ranking OT-compatible with A in n steps. Furthermore, all rankings OT-compatible with A belong to the search space of RCD. Finally, if the input tableau is not OT-compatible, RCD detects that, in the sense that it gets stucked before all constraints are ranked.⁴ I will come back to RCD in chapter 5 and note that it can be reinterpreted

⁴ The definition of RCD given in (373) is slightly different from the original definition by Tesar and Smolensky. The difference between Tesar and Smolensky's original definition of RCD and the one given here shows up for comparative tableaux that have more than one constraint with no undeleted L's. According to the definition (373), RCD arbitrarily chooses *one*

as the old Fourier-Motzkin Elimination Algorithm for polyhedral feasibility.⁵

- (373) For $t = 1$ to n :
- a. assign to the t th stratum⁶ a yet unstroked constraint whose column in \mathbf{A} does not contain any unstroked L;
 - b. strike out every row of \mathbf{A} that has a W under the constraint just picked in step (a) and then strike out the entire column corresponding to that constraint.

RCD repeats n iterations each of which takes at most nm time (the algorithm might need to scan n columns with m entries each). Thus, RCD constitutes a polynomial-time solution algorithm for the Ranking problem (371), in the sense of (354). This shows the tractability of the Ranking problem (371) and thus also of the Ranking problem (370), thus concluding the proof of claim 35.

7.4.2 Complexity of Prince and Tesar’s Subset problem

Let me restate the Subset problem (349) in terms of comparative tableaux, as in the formulation (374). The set \mathcal{F} provided with an instance of the problem says which one of the n columns of the input comparative tableau \mathbf{A} correspond to faithfulness constraints. For completeness, I have also made explicit the size of an instance of the problem, which of course depends on the two dimensions of the comparative tableau.

- (374) *given:* a) an OT-compatible input tableau $\mathbf{A} \in \{L, E, W\}^{m \times n}$;
 b) the set $\mathcal{F} \subseteq \{1, \dots, n\}$ of faithfulness constraints;
- output:* a ranking \gg OT-compatible with \mathbf{A} such that there exists no ranking \gg' OT-compatible with \mathbf{A} too such that $\mu(\gg')$ is smaller than $\mu(\gg)$, where μ is defined as in (352);
- size:* $\max\{m, n\}$.

The goal of this subsection is to prove the following claim 36, that says the Subset problem (374) is not “easy”, contrary to the Ranking problem. Prince and Tesar’s formulation (374) of the Subset problem has been very influential in the literature, and thus claim 36 is interesting in its own right. Furthermore, it will allow me to straightforwardly derive the intractability of the Subset problem (347) in the next subsection. The proof of claim 36 presented in this section works even if we restrict ourselves to instances of the problem (374) whose comparative tableau \mathbf{A} has a very simple disjunctive structure, namely contains no more than two entries equal to W per row;⁷ and even if we furthermore restrict ourselves to instances of the problem whose comparative tableau \mathbf{A} does not contain a single entry equal to L in the columns corresponding to faithfulness constraints \mathcal{F} .

Claim 36 Under the conjecture that $\mathcal{P} \neq \mathcal{NP}$, the PTSubset problem (374) is *not* tractable.

such constraint and assigns it to the highest available stratum. According to Tesar and Smolensky’s original definition, RCD assigns *all* such constraints to the highest available stratum and then outputs a total ranking which is an arbitrary refinement of the non-total ranking thus constructed. To illustrate how the two definitions differ, consider the comparative tableau in (i).

$$(i) \quad \begin{array}{c} C_1 \quad C_2 \quad C_3 \\ \left[\begin{array}{ccc} W & E & L \\ E & W & W \end{array} \right] \end{array}$$

Tesar and Smolensky’s original RCD first computes the non-total ranking $\{C_1, C_2\} \gg \{C_3\}$ and then outputs one of its two refinements, namely either $C_1 \gg C_2 \gg C_3$ or $C_2 \gg C_1 \gg C_3$. Thus, the ranking $C_1 \gg C_3 \gg C_2$ lies outside of the search space of Tesar and Smolensky’s RCD, despite the fact that this ranking too is OT-compatible with the given comparative tableau (i). Instead, the version of RCD defined in (373) might output such a ranking, provided that the algorithm chooses C_1 at the first step, C_3 at the next step and C_2 at the last step. More generally, the version of RCD defined in (373) poses no artificial restrictions on the search space, which indeed contains any ranking OT-compatible with the given comparative tableau.

⁵See Eisner (2000) for a reinterpretation of RCD as a generalization of Topological sort to directed hypergraphs.

⁶With the understanding that the 1st stratum is the top stratum and the n th stratum is the bottom stratum.

⁷Of course, if the input comparative tableau \mathbf{A} has a unique entry equal to W per row, then it is OT-compatible with a unique ranking, and thus the corresponding instance of the PTSubset problem (374) reduces to the Ranking problem, and is therefore easy.

Let me start by noting that the statement of the problem (374) can be simplified slightly. Given a ranking \gg of the constraint set $\mathcal{C} = \{C_1, \dots, C_n\}$, let me represent it with the permutation $\pi : \{C_1, \dots, C_n\} \rightarrow \{1, \dots, n\}$ such that $\pi(C_k)$ is the stratum to which the constraint C_k is assigned by \gg , with the understanding that the top stratum is the n th stratum. Thus, $C_k \gg C_h$ iff $\pi(C_k) > \pi(C_h)$. I will denote by \mathcal{S}_n the set of all such permutations. By virtue of this correspondence between a ranking \gg and a permutation π , all the notions pertaining to rankings introduced so far straightforwardly extend to permutations. Thus, I can speak of a given permutation $\pi \in \mathcal{S}_n$ being OT-compatible with a comparative tableau \mathbf{A} according to (13a). And I can think of the strictness measure μ in (352) as being defined over permutations $\pi \in \mathcal{S}_n$, obviously as in (375).

$$(375) \quad \mu(\pi) = \sum_{F \in \mathcal{F}} \left| \{M \in \mathcal{M} \mid \pi(F) > \pi(M)\} \right|$$

Once restated as in (375) in terms of permutations, Prince and Tesar's strictness measure can be equivalently described as in (376). In words, the strictness of a ranking π coincides with the sum of the strata to which π assigns the faithfulness constraints, apart from an additive constant that does not depend on the specific ranking π considered.⁸

$$(376) \quad \mu(\pi) = \sum_{F \in \mathcal{F}} \pi(F) - \text{constant}$$

The constant term that appears in (376) can of course be ignored. Thus, the PTsubset problem (374) can be restated as the equivalent problem (377). Since the two problems are equivalent by (376), I will concentrate on showing that the latter problem (377) is not tractable.

- (377) *given:* a) an OT-compatible input tableau \mathbf{A} ;
 b) the set $\mathcal{F} \subseteq \{1, \dots, n\}$ of faithfulness constraints
output: a ranking $\pi \in \mathcal{S}_n$ OT-compatible with \mathbf{A} such that there exists no ranking π' OT-compatible with \mathbf{A} too such that $\sum_{F \in \mathcal{F}} \pi'(F) < \sum_{F \in \mathcal{F}} \pi(F)$.

The decision problem corresponding to problem (377) is (378). From now on, I will refer to (378) as the PTSUBSET PROBLEM. As noted for the general case in section 7.3, the non-tractability of the decision problem (378) entails the non-tractability of the original problem (377). In fact, if the original problem (377) can be solved in polynomial time, then the corresponding decision problem (378) can be solved in polynomial time too: given an instance of the decision problem (378), find a solution π of the corresponding instance of the problem (377) and then just check whether $\sum_{F \in \mathcal{F}} \pi(F) \leq k$.

⁸The identity in (376) is obvious. For completeness, I prove it explicitly in this footnote. Let $\ell \leq n$ be the total number of faithfulness constraints F_1, \dots, F_ℓ . Let $j_k \in \{1, \dots, n\}$ be the stratum to which the ranking π assigns the faithfulness constraint F_k , namely $j_k = \pi(F_k)$ for $k = 1, \dots, \ell$. Assume without loss of generality that $j_1 < j_2 < \dots < j_\ell$ (otherwise, just relabel the constraints). Thus, the lowest stratum to which π assigns a faithfulness constraint is the j_1 th stratum, the next lowest stratum to which π assigns a faithfulness constraint is the j_2 th stratum, and so on. The proof of the identity (376) consists of the chain of identities in (i). In step (a), I have used the definition (375) of the strictness measure μ . In step (b), I have reasoned as follows: the number of markedness constraints ranked by π below the faithfulness constraint F_k is equal to the total number of constraints ranked by π below F_k (namely $j_k - 1$, since F_k is assigned to the j_k th stratum) minus the number of faithfulness constraints ranked by π below F_k (which is $k - 1$, since F_k is the k th faithfulness constraint starting from the bottom of the ranking).

$$\begin{aligned} \text{(i)} \quad \mu(\pi) &\stackrel{\text{(a)}}{=} \sum_{k=1}^{\ell} \left| \{M \in \mathcal{M} \mid \pi(F_k) > \pi(M)\} \right| \\ &\stackrel{\text{(b)}}{=} \sum_{k=1}^{\ell} (j_k - k) \\ &= \sum_{k=1}^{\ell} j_k - \sum_{k=1}^{\ell} k \\ &= \sum_{k=1}^{\ell} \pi(F_k) - \text{constant} \end{aligned}$$

Prince and Tesar (2004) stick to the slightly more complicated definition (375) because they consider nontotal hierarchies. In fact, the equivalence in (376) only holds for total hierarchies, as it is obvious from the proof (i).

- (378) *given:* a) an OT-compatible input tableau A with n columns;
 b) the set $\mathcal{F} \subseteq \{1, \dots, n\}$ of faithfulness constraints;
 c) an integer k ;
output: “yes” iff there exists a ranking $\pi \in \mathcal{S}_n$ that is OT-compatible with A and furthermore such that $\sum_{F \in \mathcal{F}} \pi(F) \leq k$.

By (368), in order to prove claim 36, I only need to exhibit a decision problem Π such that Π is NP-complete and furthermore $\Pi \leq_P$ PTSUBSET. Let me introduce the decision problem Π that I will use to this end. Given an arbitrary finite set $A = \{a, b, \dots\}$ with cardinality $|A|$, consider a set S of pairs of elements of A . The set S is called LINEARLY COMPATIBLE iff there exists a one-to-one function $\pi : A \rightarrow \{1, 2, \dots, |A|\}$ such that for every pair $(a, b) \in S$ we have $\pi(a) < \pi(b)$. This notion is illustrated in (379): the set S in (379a) is linearly compatible; the one in (379b) is not.

- (379) $A = \{a, b, c, d\}$
 a. $S = \{(a, b), (b, c), (c, d)\}$
 b. $S = \{(a, b), (b, c), (c, a)\}$

It is useful to let S be not just a set but a MULTISSET, namely to allow for the possibility that S contains multiple instances of the same pair. The notion of cardinality and the subset relation are trivially extended from sets to multisets. Thus, consider the decision problem (380), that I will call the MAX-ORDERING PROBLEM. This problem is obviously in \mathcal{NP} , namely it admits a verification algorithm in the sense of (363).

- (380) *given:* a) a finite set A ;
 b) a multiset $P \subseteq A \times A$ of pairs of the elements of A ;
 c) an integer $k \leq |P|$;
output: “yes” iff there exists a multiset $S \subseteq P$ with $|S| \geq k$ linearly compatible;
size: $\max\{|A|, |P|\}$.

The next claim 37 ensures that the MaxOrdering problem is NP-complete. The following claim 38 shows that MaxOrdering \leq_P PTSUBSET. By (368), I can thus conclude that the problem (378) is non-tractable, thus completing the proof of claim 36.

Claim 37 The MaxOrdering problem is NP-complete.⁹

Proof. By (367), in order to prove that MaxOrdering is NP-complete, I need to prove (381a). To this end, it is sufficient to prove (381b). In fact, the NP-completeness of Π in (381b) ensures that $\Pi' \leq_P \Pi$ for every decision problem $\Pi' \in \mathcal{NP}$; this fact together with the transitivity of the relation \leq_P entail that also $\Pi' \leq_P$ MaxOrdering for every decision problem $\Pi' \in \mathcal{NP}$, namely that (381a) holds.

- (381) a. For every decision problem $\Pi' \in \mathcal{NP}$:
 $\Pi' \leq_P$ MaxOrdering.
 b. For some NP-complete problem Π :
 $\Pi \leq_P$ MaxOrdering.

Let me introduce the decision problem Π that I will use for proving (381b). Given an arbitrary finite set $A = \{a, b, \dots\}$ with cardinality $|A|$, consider a set T of triplets of elements of A . The set T is called LINEARLY CYCLICALLY COMPATIBLE iff there exists a one-to-one function $\pi : A \rightarrow \{1, 2, \dots, |A|\}$ such that for every triplet $(a, b, c) \in T$ either $\pi(a) < \pi(b) < \pi(c)$ or $\pi(b) < \pi(c) < \pi(a)$ or $\pi(c) < \pi(a) < \pi(b)$. This notion is illustrated in (382): the set T in (382a) is linearly cyclically compatible; the one in (382b) is not.

- (382) $A = \{a, b, c, d\}$
 a. $T = \{(a, b, c), (b, c, d)\}$

⁹The MaxOrdering problem looks general enough for people to have proven its NP-completeness already. Yet, I couldn't find it in the literature, but for a rather similar claim in Cohen et al. (1999).

$$b. T = \{(a, b, c), (a, c, b)\}$$

Consider the CYCLICORDERING PROBLEM in (383).¹⁰ Its NP-completeness was proven by reduction from 3SATISFABILITY in Galil and Megiddo (1977); CyclicOrdering is problem [MS2] in Garey and Johnson (1979, p. 279).

- (383) *input:* a) a finite set A ;
 b) a collection $T \subseteq A \times A \times A$ of triplets of elements of A ;
output: “yes” iff T is linearly cyclically compatible;
size: the cardinality $|A|$ of A .

Given an instance instance (A, T) of CyclicOrdering, consider the corresponding instance (A, P, k) of MaxOrdering defined as in (384). For every triplet (a, b, c) in the set T , we put in the multiset P the three pairs (a, b) , (b, c) and (c, a) . Furthermore, we set the threshold k to twice the number of triplets in the set T .

$$(384) \quad P = \{(a, b), (b, c), (c, a) \mid (a, b, c) \in T\}$$

$$k = 2|T|$$

The construction is illustrated in (385): given the instance of CyclicOrdering in (385a), we construct the corresponding instance of MaxOrdering in (385b). Note that P is a multiset because it contains two instances of the pair (b, c) , coming from two different triplets in T .

$$(385) \quad a. \left(A = \{a, b, c, d\}, \quad T = \{(a, b, c), (b, c, d)\} \right)$$

$$b. \left(A = \{a, b, c, d\}, \quad P = \left\{ \begin{array}{l} (a, b), (b, c), (c, a), \\ (b, c), (c, d), (d, b) \end{array} \right\}, \quad k = 4 \right)$$

In order to show that CyclicOrdering \leq_P MaxOrdering, let me show that the correspondence defined in (384) is a reduction algorithm according to (359). This correspondence is trivially computable in time polynomial in $|A|$. Thus, I only need to show that an instance (A, T) of CyclicOrdering admits a positive answer iff the corresponding instance (A, P, k) of MaxOrdering admits a positive answer. If the instance (A, T) of CyclicOrdering admits a positive answer, then there exists a linear order π on A cyclically compatible with T ; this means in turn that for every triplet $(a, b, c) \in T$, there are at least two pairs in P compatible with π and thus there are a total of $k = 2|T|$ pairs in P compatible with π .¹¹ Vice versa, if the instance (A, P, k) of MaxOrdering admits a positive answer, then there exists a linear order π on A compatible with $2|T|$ pairs in P ; since the three pairs that come from a given triplet are inconsistent, then each triplet must contribute two pairs to the total of $2|T|$ compatible pairs and thus π must be cyclically compatible with all the triplets in T . ■

Claim 38 MaxOrdering \leq_P PTSubset.

Proof. Given an instance (A, P, k) of the MaxOrdering problem (380), let $n = |A|$ and $\ell = |P|$; pick an integer $d \in \mathbb{N}$ that satisfies (386).

$$(386) \quad d > n(\ell - k) + \frac{1}{2}\ell^2 + \frac{1}{2}\ell$$

Consider the corresponding instance (A, \mathcal{F}, K) of the PTSubset problem (378) defined as follows. Let the numbers N and M of columns and of rows of the tableau A and the threshold K be defined as in (387).

¹⁰It makes sense to let the size of an instance of the CyclicOrdering problem (383) be just the cardinality of the set A . In fact, the cardinality of the set T can be at most $|A|^3$. On the other hand, it makes sense to let the size of an instance of the MaxOrdering problem (380) depend also on the cardinality of the multiset P rather than only on the cardinality of the set A , as in the case of the CyclicOrdering problem (383). In fact, being P a multiset, its cardinality cannot be bound by the cardinality of A .

¹¹Note that, in order for the latter claim to hold, it is crucial that P be a multiset, namely that the same pair might be counted twice. In fact, T might contain two different triplets that share some elements, such as (a, b, c) and (a, b, d) .

$$(387) \quad \begin{aligned} N &= \ell + n + d \\ M &= \ell + nd \\ K &= (\ell - k)(n + d) + \frac{1}{2}\ell^2 + \frac{1}{2}\ell \end{aligned}$$

The sets \mathcal{F} and \mathcal{M} of faithfulness and markedness constraints are defined as in (388). There is a faithfulness constraint F_1, \dots, F_ℓ for every pair in the multiset P given with the instance of Max-Ordering. Markedness constraints come in two varieties, annotated with and without the prime. In particular, there is a markedness constraint M_1, \dots, M_n for every element in the set A given with the instance of MaxOrdering.

$$(388) \quad \begin{aligned} \mathcal{F} &= \{F_1, \dots, F_\ell\} \\ \mathcal{M} &= \{M_1, \dots, M_n\} \cup \{M'_1, \dots, M'_d\} \end{aligned}$$

The comparative tableau $\mathbf{A} \in \{L, E, W\}^{M \times N}$ is built in two steps. First, the smaller comparative tableau $\overline{\mathbf{A}} \in \{L, E, W\}^{\ell \times (\ell + n)}$ is built as in (389). This small comparative tableau $\overline{\mathbf{A}}$ has a row for every pair $(a_i, a_j) \in P$. This comparative row contains all E's but for three entries: the entry corresponding to the faithfulness constraint $F_{(i,j)}$ corresponding to that pair, which is W; the entry corresponding to the markedness constraint M_i corresponding to the first element a_i in the pair, which is L; the entry corresponding to the markedness constraint M_j corresponding to the second element a_j in the pair, which is W. The intuition is that a linear order σ over A is compatible with a given pair $(a_i, a_j) \in P$ iff the entry L in the corresponding row of the comparative tableau $\overline{\mathbf{A}}$ is accounted for by ranking M_j over M_i without any need for the corresponding faithfulness constraint $F_{(i,j)}$ to do any work.

$$(389) \quad (a_i, a_j) \in P \Rightarrow \left[\begin{array}{cccc|cccc} \dots & F_{(i,j)} & \dots & \dots & M_i & \dots & M_j & \dots \\ & \vdots & & & \vdots & & \vdots & \\ \dots & W & \dots & \dots & L & \dots & W & \dots \\ & \vdots & & & \vdots & & \vdots & \end{array} \right] = \overline{\mathbf{A}}$$

I illustrate this construction in (390): given the set A and the multiset of pairs P in (390a), the corresponding small comparative tableau $\overline{\mathbf{A}}$ is (390b).

$$(390) \quad \begin{aligned} \text{a. } & A = \{a, b, c\}, P = \{(a, b), (b, c), (c, a)\} \\ \text{b. } & \begin{array}{l} (a,b) \\ (b,c) \\ (c,a) \end{array} \left[\begin{array}{cccccc} F_{(a,b)} & F_{(b,c)} & F_{(c,a)} & M_a & M_b & M_c \\ W & & & L & W & \\ & W & & & L & W \\ & & W & W & & L \end{array} \right] = \overline{\mathbf{A}} \end{aligned}$$

Suppose that an L corresponding to a markedness constraint M_i in the small comparative tableau $\overline{\mathbf{A}}$ in (389) is not accounted for by high ranking the corresponding markedness constraint M_j so that the corresponding faithfulness constraint $F_{(i,j)}$ needs to be high ranked instead. What consequences has this fact for the overall strictness measure (376)? Not much: all I can deduce is that the faithfulness constraint $F_{(i,j)}$ has at least the two markedness constraints M_i and M_j ranked below it. To get a more dramatic effect, I adopt the following simple trick: I add a bunch of new markedness constraints M'_1, \dots, M'_d and force them to be ranked below every markedness constraint M_1, \dots, M_n . Thus, if the faithfulness constraint $F_{(i,j)}$ is ranked above M_i and M_j , then it must also be ranked above all these extra markedness constraints M'_1, \dots, M'_d . If the number d of these extra constraints is properly chosen, as in (386), then the corresponding effect on the strictness measure (376) is rather dramatic. Here are the details. Embed the small comparative tableau $\overline{\mathbf{A}} \in \{L, E, W\}^{\ell \times (\ell + n)}$ into the larger comparative tableau $\mathbf{A} \in \{L, E, W\}^{M \times N}$ as in (391). The first horizontal block of the comparative tableau \mathbf{A} basically is just a copy of the small comparative tableau $\overline{\mathbf{A}}$, with many E's underneath the new markedness constraints M'_1, \dots, M'_d . This first horizontal block is followed by n new horizontal blocks. For every $i = 1, \dots, n$, the effect of the i th such block is to force all the new constraints M'_1, \dots, M'_d to be ranked underneath the markedness constraint M_i .

$$(391) \quad \left[\begin{array}{c} \lceil \\ \vdots \\ \lceil \\ \text{first block of } d \text{ rows} \\ \vdots \\ \lceil \\ \text{nth block of } d \text{ rows} \\ \vdots \\ \lceil \\ \vdots \\ \lceil \end{array} \right] \left[\begin{array}{ccc|ccc} F_1 & \dots & F_\ell & M_1 & \dots & M_n & M'_1 & \dots & M'_d \\ \hline & & \overline{A} & & & & & & \\ \hline & & & W & & & L & & \\ & & & | & & & \diagdown & & \\ & & & W & & & & & L \\ \hline & & \vdots & & & & \vdots & & \\ \hline & & & & & W & L & & \\ & & & & & | & \diagdown & & \\ & & & & & W & & & L \end{array} \right] \doteq A$$

In order to show that $\text{MaxOrdering} \leq_P \text{PTSubset}$, let me show that the correspondence just defined is a reduction algorithm according to (359). Clearly, this reduction can be computed in polynomial time. In the rest of the proof, I show that the instance (A, P, k) of MaxOrdering admits a positive answer iff the corresponding instance (A, \mathcal{F}, K) of PTSubset admits a positive answer. Let me start by assuming that the instance (A, P, k) of MaxOrdering admits a positive answer. Thus, there exists a sub-multiset $S \subseteq P$ of cardinality k compatible with a linear order $\sigma \in \mathcal{S}_n$ on A . To simplify the notation, assume without loss of generality that the k pairs in S are the first k pairs in P . Consider a permutation $\pi \in \mathcal{S}_N$ that corresponds to a ranking \gg that satisfies (392): π assigns the k faithfulness constraints that correspond to pairs in S to the k bottom strata of the hierarchy in any order; π assigns the d special markedness constraints M'_1, \dots, M'_d to the next d strata; π assigns the n markedness constraints M_1, \dots, M_n to the next n strata ordered according to σ ; finally, π assigns the remaining $\ell - k$ faithfulness constraints to the top $\ell - k$ strata in any order.

$$(392) \quad \{F_{k+1}, \dots, F_\ell\} \gg M_{\sigma(n)} \gg \dots \gg M_{\sigma(1)} \gg \{M'_1, \dots, M'_d\} \gg \{F_1, \dots, F_k\}$$

This permutation $\pi \in \mathcal{S}_N$ is OT-compatible with the comparative tableau A in (391). In fact, it is OT-compatible with the bottom n horizontal blocks, since M_1, \dots, M_n are ranked above M'_1, \dots, M'_d . It is OT-compatible with the bottom $\ell - k$ rows of the small comparative tableau \overline{A} , since F_{k+1}, \dots, F_ℓ are ranked at the top. Finally, it is OT-compatible with the top k rows of the small comparative tableau \overline{A} . In fact, assume by contradiction that it were not. This would mean that there exists one such row such that π is not OT-compatible with that row; this means in turn that that row has a w under the column M_j , a L under the column M_i and $\pi(M_j) < \pi(M_i)$; this means in turn that $\sigma(a_j) < \sigma(a_i)$ and thus σ is not compatible with the pair (a_i, a_j) contradicting the hypothesis that that pair belongs to S . Finally, the chain of inequalities in (393) shows that $\sum_{F \in \mathcal{F}} \pi(F) \leq K$ and thus that the corresponding instance of PTSubset admits a positive answer.

$$(393) \quad \begin{aligned} \sum_{F \in \mathcal{F}} \pi(F) &= \\ &= \sum_{h=1}^k \pi(F_h) + \sum_{h=k+1}^{\ell} \pi(F_h) = \\ &= (1 + 2 + \dots + k) + ((k + n + d + 1) + (k + n + d + 2) + \dots + (k + n + d + (\ell - k))) \\ &= \frac{1}{2}k(k + 1) + (k + n + d)(\ell - k) + \frac{1}{2}(\ell - k)(\ell - k + 1) \\ &= \frac{1}{2}k^2 + \frac{1}{2}k + kl + nl + dl - k^2 - kn - kd + \frac{1}{2}\ell^2 - \frac{1}{2}k\ell - \frac{1}{2}k\ell + \frac{1}{2}k^2 + \frac{1}{2}\ell - \frac{1}{2}k \\ &= nl + dl - kn - kd + \frac{1}{2}\ell^2 + \frac{1}{2}\ell \\ &= K \end{aligned}$$

Vice versa, assume now that the instance (A, \mathcal{F}, K) of PTSubset admits a positive answer. This means that there exists a permutation $\pi \in \mathcal{S}_N$ OT-compatible with the comparative tableau A such that $\sum_{F \in \mathcal{F}} \pi(F) \leq K$. Let σ be the linear order on $A = \{a_1, \dots, a_n\}$ defined by $\sigma(a_i) > \sigma(a_j)$ iff $\pi(M_i) > \pi(M_j)$. Consider the multiset $S \subseteq P$ in (394).

$$(394) \quad S = \left\{ (a_h, a_k) \in P \mid \pi(M_h) < \pi(M_k) \right\}$$

Clearly, S is compatible with the linear order σ . To prove that the instance (A, P, k) of MaxOrdering has a positive answer, it is thus sufficient to show that $|S| \geq k$. Assume by contradiction that $|S| = x < k$. This means that there are $\ell - x$ rows in the small comparative tableau \bar{A} such that the markedness constraint with a w in that row is π -ranked below the markedness constraint with a L in that row; hence, in order for π to be OT-compatible with that row, it must rank the faithfulness constraint corresponding to that row above the two markedness constraints involved in that row and thus also above all the d extra markedness constraints M'_1, \dots, M'_d . This means in turn that each of these $\ell - x$ faithfulness constraints has at least d markedness constraints ranked below. The chain of inequalities in (395) thus holds. Here, I have reasoned as follows: in step (a), I have used the definition (387) of the constant K ; in step (b), I have used the hypothesis that π verifies the given instance of the PTSubset problem; in step (c), I have used the fact just noted that there are at least $\ell - x$ faithfulness constraints $F \in \mathcal{F}$ such that $\pi(F) \geq d$; in step (d), I have used the contradictory assumption that $x \leq k - 1$.

$$(395) \quad (\ell - k)d + (\ell - k)n + \frac{1}{2}\ell^2 + \frac{1}{\ell}\ell \stackrel{(a)}{=} K \\ \stackrel{(b)}{\geq} \sum_{F \in \mathcal{F}} \pi(F) \\ \stackrel{(c)}{\geq} (\ell - x)d \\ \stackrel{(d)}{\geq} (\ell - (k - 1))d \\ = (\ell - k)d + d$$

But the inequality $d \leq (\ell - k)n + \frac{1}{2}\ell^2 + \frac{1}{\ell}\ell$ thus derived in (395) contradicts the choice (386) of the constant d . ■

7.4.3 Complexity of the Subset problem

Let me now turn to the original formulation of the Subset problem (347), repeated in (396). The problem could in principle be restated by replacing data sets with comparative tableaux, as done in the preceding subsection for the PTSubset problem. Yet, in the case of the Subset problem, I would still need to provide, along with the comparative tableau, also the universal specifications \mathcal{X} , \mathcal{Y} , Gen and \mathcal{C} , in order to be able to compute the language $\mathcal{R}(\text{OT}_{\gg})$ corresponding to a ranking \gg . And I would also need to make sure that the universal specifications and the comparative tableau given with an instance of the problem do match each other. For this reason, it is more convenient to stop at the formulation (396), and avoid restating it in terms of comparative tableaux. In a sense, Prince and Tesar's (2004) formulation (349) of the problem in terms of strictness measures can indeed be seen as a way of formulating the Subset problem in terms of comparative tableaux, thus bringing it closer to the successful formulation given by Tesar and Smolensky for the Ranking problem.

$$(396) \quad \text{given:} \quad \begin{array}{l} \text{a) universal specifications } \mathcal{X}, \mathcal{Y}, Gen \text{ and } \mathcal{C}; \\ \text{b) a finite OT-compatible data set } \mathcal{D} \subseteq \mathcal{X} \times \mathcal{Y}; \end{array} \\ \text{output:} \quad \text{a ranking } \gg \text{ OT-compatible with } \mathcal{D} \text{ such that there exists no other ranking} \\ \gg' \text{ OT-compatible with } \mathcal{D} \text{ too such that } \mathcal{R}(\text{OT}_{\gg'}) \subsetneq \mathcal{R}(\text{OT}_{\gg}).$$

The issue just discussed is strictly connected with the issue of the proper definition of the size of an instance of the Subset problem (396). In (370), I have let the size of an instance of the Ranking problem depend on the complexity of the data set \mathcal{D} , measured in terms of its cardinality $|\mathcal{D}|$ and in terms of the maximum number $width(\mathcal{D})$ of candidates that an underlying form in \mathcal{D} is paired with, as defined in (369). But in the case of the Subset problem (396) it makes sense to let the complexity of an instance of the problem depend on the complexity of the universal specifications, that of course upper bounds the complexity of any corresponding data set. Thus, I complete the definition of the problem as in (397), whereby the size of an instance of the problem depends on the complexity

of the universal specifications, measured in terms of the number $|\mathcal{X}|$ of underlying forms and the maximum number $width(\mathcal{X})$ of candidates that an underlying form is paired with. From now on, I will refer to (397) as the SUBSET PROBLEM.

- (397) *given:* a) universal specifications $\mathcal{X}, \mathcal{Y}, Gen$ and \mathcal{C} ;
 b) a finite OT-compatible data set $\mathcal{D} \subseteq \mathcal{X} \times \mathcal{Y}$;
output: a ranking \gg OT-compatible with \mathcal{D} such that there exists no other ranking \gg' OT-compatible with \mathcal{D} too such that $\mathcal{R}(OT_{\gg'}) \subsetneq \mathcal{R}(OT_{\gg})$.
size: $\max\{|\mathcal{C}|, |\mathcal{X}|, width(\mathcal{X})\}$

The goal of this section is to prove the following claim 39, that says the Subset problem (397) is not “easy”, contrary to the Ranking problem. This claim says in particular that the non-tractability of Prince and Tesar’s (2004) formulation of the Subset problem is not an idiosyncratic property of that specific formulation but plausibly reflects the intrinsic complexity of the problem.

Claim 39 Under the conjecture that $\mathcal{P} \neq \mathcal{NP}$, the Subset problem (397) is *not* tractable.

As usual, consider the decision variant (398) of the Subset problem (397). Once again, in order to show that problem (397) is not tractable, it is sufficient to show that the corresponding decision problem (398) is not tractable. In fact, if (397) can be solved in polynomial time, then (398) can be solved in polynomial time too: given an instance of the decision problem (398), find a solution \gg of the corresponding instance of the problem (397) and then just check whether $|\mathcal{R}(OT_{\gg})| \leq k$ (note that the definition of the size of the problem allows us to check in polynomial time whether $|\mathcal{R}(OT_{\gg})| \leq k$ for any given ranking \gg).

- (398) *given:* a) universal specifications $\mathcal{X}, \mathcal{Y}, Gen$ and \mathcal{C} ;
 b) a finite OT-compatible data set $\mathcal{D} \subseteq \mathcal{X} \times \mathcal{Y}$;
 c) an integer k ;
output: “yes” iff there exists a ranking \gg OT-compatible with \mathcal{D} such that the corresponding language $\mathcal{R}(OT_{\gg})$ contains at most k surface forms;
size: $\max\{|\mathcal{C}|, |\mathcal{X}|, width(\mathcal{X})\}$.

The following claim 40 ensures that the PTSubset problem (374) can be reduced to the Subset problem (398). Since the PTSubset problem is NP-complete by claim 36, then I conclude by (368) that the Subset problem (398) is NP-complete too, thus completing the proof of claim 39.

Claim 40 $PTsubset \leq_P Subset$.

Proof. Given an instance (A, \mathcal{F}, k) of the PTSubset problem (378), let n be the number of columns of the comparative tableau A , namely the total number of constraints; let m be the number of rows of the comparative tableau A ; let ℓ be the cardinality of the set \mathcal{F} , namely the total number of faithfulness constraints. Let the corresponding instance $(\mathcal{X}, \mathcal{Y}, Gen, \mathcal{C}, \mathcal{D}, K)$ of the Subset problem (398) be defined as follows. Define the threshold K , the set \mathcal{X} of underlying forms, the set \mathcal{Y} of surface forms, the generating function Gen and the data set \mathcal{D} as in (399).

- (399) a. $K = m + k + d$, where $d = \ell(n - \ell)$
 b. $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2 \cup \mathcal{X}_3$,
 where $\mathcal{X}_1 = \{x_1, \dots, x_m\}$, $\mathcal{X}_2 = \{x'_1, \dots, x'_d\}$ and $\mathcal{X}_3 = \{x''_1, \dots, x''_d\}$
 c. $\mathcal{Y} = \mathcal{Y}_1 \cup \mathcal{Y}_2 \cup \mathcal{Y}_3$,
 where $\mathcal{Y}_1 = \left\{ \begin{array}{l} y_1, \dots, y_m, \\ z_1, \dots, z_m \end{array} \right\}$, $\mathcal{Y}_2 = \left\{ \begin{array}{l} u_1, \dots, u_d, \\ v_1, \dots, v_d \end{array} \right\}$ and $\mathcal{Y}_3 = \left\{ \begin{array}{l} u_1, \dots, u_d, \\ w_1, \dots, w_d \end{array} \right\}$
 d. $Gen(x_i) = \{y_i, z_i\} \subseteq \mathcal{Y}_1$ for every $x_i \in \mathcal{X}_1$
 $Gen(x'_i) = \{u_i, v_i\} \subseteq \mathcal{Y}_2$ for every $x'_i \in \mathcal{X}_2$
 $Gen(x''_i) = \{u_i, w_i\} \subseteq \mathcal{Y}_3$ for every $x''_i \in \mathcal{X}_3$

$$e. \mathcal{D} = \{(x_1, y_1), \dots, (x_m, y_m)\} \subseteq \mathcal{X}_1 \times \mathcal{Y}_1$$

Let the constraint set \mathcal{C} contain a total of n constraints C_1, \dots, C_n ; let C_k be a faithfulness constraint iff $k \in \mathcal{F}$; let C_k be a markedness constraint otherwise. Define these n constraints separately on $\mathcal{X}_1 \times \mathcal{Y}_1$, $\mathcal{X}_2 \times \mathcal{Y}_2$ and $\mathcal{X}_3 \times \mathcal{Y}_3$ as follows (there is no need to define the constraints on $\mathcal{X}_i \times \mathcal{Y}_j$ for $i \neq j$, because of the definition of the *Gen* function). The set \mathcal{X}_1 contains a total of m underlying forms x_1, \dots, x_m , one for every row of the given comparative tableau **A**. Each of these underlying forms $x_i \in \mathcal{X}_1$ comes with two candidates y_i and z_i that together make up the entire set of surface forms \mathcal{Y}_1 . Define the constraints C_1, \dots, C_n over $\mathcal{X}_1 \times \mathcal{Y}_1$ according to the condition (400). This condition ensures that the comparative tableau **A** _{\mathcal{D}} corresponding to the set of data \mathcal{D} in (399e) is indeed the given comparative tableau **A**.

$$(400) \quad \begin{array}{ll} C_k(x_i, y_i) < C_k(x_i, z_i) & \iff \text{the } k\text{th entry in the } i\text{th row of } \mathbf{A} \text{ is a w} \\ C_k(x_i, y_i) = C_k(x_i, z_i) & \iff \text{the } k\text{th entry in the } i\text{th row of } \mathbf{A} \text{ is a E} \\ C_k(x_i, y_i) > C_k(x_i, z_i) & \iff \text{the } k\text{th entry in the } i\text{th row of } \mathbf{A} \text{ is a L} \end{array}$$

The set \mathcal{X}_2 contains a total of $d = \ell(n - \ell)$ underlying forms x'_1, \dots, x'_d , one for every pair of a faithfulness constraint and a markedness constraint. Pair up (in some arbitrary but fixed way) each underlying form $x'_i \in \mathcal{X}_2$ with a unique pair of a faithfulness constraint and a markedness constraint; thus, I can speak of the markedness and the faithfulness constraints “corresponding” to a given underlying form $x'_i \in \mathcal{X}_2$. Each of these underlying forms $x'_i, \dots, x'_d \in \mathcal{X}_2$ comes with two candidates u_i and v_i that together make up the entire set of surface forms \mathcal{Y}_2 . Define the constraints C_1, \dots, C_n over $\mathcal{X}_2 \times \mathcal{Y}_2$ according to the condition (401). This condition ensures that the grammar OT_{\gg} corresponding to an arbitrary ranking \gg maps the underlying form x'_i to the surface form v_i iff the faithfulness constraint corresponding to the underlying form x'_i is ranked \gg -higher than the faithfulness constraint corresponding to x'_i .

$$(401) \quad \begin{array}{ll} C_k(x'_i, v_i) < C_k(x'_i, u_i) & \text{if } C_k \text{ is the faithfulness constraint corresponding to } x'_i \\ C_k(x'_i, v_i) > C_k(x'_i, u_i) & \text{if } C_k \text{ is the markedness constraint corresponding to } x'_i \\ C_k(x'_i, v_i) = C_k(x'_i, u_i) & \text{otherwise} \end{array}$$

Finally, define the constraints C_1, \dots, C_n over $\mathcal{X}_3 \times \mathcal{Y}_3$ according to the condition (402). This condition ensures for any ranking \gg that the corresponding language $\mathcal{R}(\text{OT}_{\gg})$ contains the surface forms u_1, \dots, u_d , namely that these forms are unmarked (as the forms [ta] and [rat] in the typology (1) considered at the beginning of the chapter).

$$(402) \quad C_k(x'_i, u_i) \leq C_k(x'_i, w_i) \quad \text{for every constraint } C_k$$

In order to show that $\text{PTSubset} \leq_P \text{Subset}$, let me show that the correspondence just defined is a reduction algorithm according to (359). Clearly, this reduction can be computed in polynomial time. In the rest of the proof, I show that an instance $(\mathbf{A}, \mathcal{F}, k)$ of the PTSubset problem (378) admits a positive answer iff the corresponding instance $(\mathcal{X}, \mathcal{Y}, \text{Gen}, \mathcal{C}, \mathcal{D}, K)$ of the Subset problem (398) admits a positive answer. If the instance $(\mathbf{A}, \mathcal{F}, k)$ of the PTSubset problem admits a positive answer, then there exists a ranking \gg OT-compatible with the comparative tableau **A** such that there are at most k pairs of a faithfulness constraint and a markedness constraint such that that faithfulness constraint is \gg -ranked above that markedness constraint. Since \gg is OT-compatible with **A** and since **A** is by construction the comparative tableau corresponding to the data set \mathcal{D} , then \gg is OT-compatible with the data set \mathcal{D} . Furthermore, the language $\mathcal{R}(\text{OT}_{\gg})$ corresponding to the ranking \gg contains at most $K = m + k + d$ surface forms, namely: the m surface forms $y_1, \dots, y_m \in \mathcal{Y}_1$ (because \gg is OT-compatible with the data set \mathcal{D}); all the d surface forms u_1, \dots, u_d (because no ranking prefers the pair (x'_i, w_i) to the pair (x'_i, u_i) by construction) and at most k of the surface forms v_1, \dots, v_d (because one of these forms v_i belongs to a given language iff the corresponding ranking ranks the faithfulness constraint corresponding to the underlying form x'_i above the corresponding markedness constraint). The vice versa holds for exactly the same reasons. ■

Chapter 8

Consequences for the relationship between standard and linear OT

In the preceding chapters, I have worked squarely within the framework of STANDARD OT, as defined by Prince and Smolensky (2004). Let me recall from section 4.1 the basic shape of this framework. Given some universal specifications $(\mathcal{X}, \mathcal{Y}, Gen, C)$, consider a ranking vector $\theta \in \mathbb{R}^n$, an underlying form $x \in \mathcal{X}$ and two corresponding candidates $y^*, y \in Gen(x)$, with the understanding that y^* is the intended winner while y is a loser candidate. We say that the ranking vector θ is OT-COMPATIBLE with the underlying/winner/loser form triplet (x, y^*, y) iff condition (403) holds. For every constraint C_k , we call the quantity $C_k(x, y) - C_k(x, y^*)$ the corresponding CONSTRAINT DIFFERENCE. It is the difference between the number $C_k(x, y)$ of violations w.r.t. constraint C_k incurred by the mapping of the underlying form x into the loser candidate y and the number $C_k(x, y^*)$ of violations w.r.t. the same constraint C_k incurred by the mapping of the underlying form x into the winner candidate y^* . As stated in (8), constraints with positive difference are called winner-preferring and constraints with negative difference are called loser-preferring. Condition (403) thus just states the usual requirement that there be at least one winner-preferring constraint that is ranked above every loser-preferring constraint.

$$(403) \quad \max_{\{h \mid C_h(x, y) - C_h(x, y^*) > 0\}} \theta_h > \max_{\{k \mid C_k(x, y) - C_k(x, y^*) < 0\}} \theta_k$$

Condition (403) is only sensitive to the sign of the constraint differences, not to their actual values. For this reason, it is useful to actually get rid of the actual values of the constraint differences, by pairing up an underlying/winner/loser form triplet (x, y^*, y) with the corresponding comparative row $\mathbf{a} = (a_1, \dots, a_k, \dots, a_n) \in \{L, E, W\}^n$ in (14), repeated in (404).

$$(404) \quad \begin{array}{c} \text{winner} \\ | \\ (x, y^*, y) \\ | \\ \text{loser} \end{array} \implies \mathbf{a} = [a_1 \quad \dots \quad a_k \quad \dots \quad a_n]$$

where $a_k \doteq \begin{cases} W & \text{if } C_k \text{ is winner-preferring} \\ L & \text{if } C_k \text{ is loser-preferring} \\ E & \text{if } C_k \text{ is inactive} \end{cases}$

Recall from (26), that $W(\mathbf{a})$ and $L(\mathbf{a})$ are the sets of winner- and loser-preferring constraints w.r.t. (the underlying/winner/loser form triplet corresponding to) the comparative row \mathbf{a} . Condition (403) can then be rewritten more compactly as in (405).

$$(405) \quad \max_{h \in W(\mathbf{a})} \theta_h > \max_{k \in L(\mathbf{a})} \theta_k$$

In this chapter, I wish to compare the framework of standard OT just reviewed with the alternative framework of LINEAR OT. This alternative framework only differs from the standard framework because of the fact that the condition (403) for compatibility between a ranking vector θ and an

underlying/winner/loser form triplet is replaced by condition (406). Let me call this alternative notion of compatibility LINEAR COMPATIBILITY (henceforth: L-compatibility).

$$(406) \quad \sum_{h=1}^n (C_k(x, y) - C_k(x, y^*)) \theta_h > 0$$

Contrary to condition (403), condition (406) does depend on the actual values of the constraint differences. For ease of comparison, let me make the notation as parallel as possible between the two frameworks and thus denote by \bar{a}_k the constraint difference corresponding to constraint C_k .

$$(407) \quad \begin{array}{c} \text{winner} \\ | \\ (x, y^*, y) \\ | \\ \text{loser} \end{array} \implies \bar{\mathbf{a}} = [\bar{a}_1 \quad \dots \quad \bar{a}_k \quad \dots \quad \bar{a}_n]$$

where $\bar{a}_k \doteq C_k(x, y) - C_k(x, y^*)$

With this piece of notation, condition (406) can be rewritten more compactly as in (408), where I am using again the scalar product $\langle \cdot, \cdot \rangle$ introduced in (94) in chapter 4.

$$(408) \quad \langle \theta, \bar{\mathbf{a}} \rangle = \sum_{h=1}^n \theta_h \bar{a}_h > 0$$

It turns out that some of the developments presented in chapter 5 have interesting consequences for the issue of the relationship between the two frameworks of standard and linear OT. I discuss some of these consequences in this final chapter.

8.1 The relationship between OT-compatibility and L-compatibility

The following claim 41 says that OT-compatibility of a comparative tableau \mathbf{A} can be characterized in terms of L-compatibility of the corresponding derived numerical matrices. This claim patterns with claims 1, 12 and 15 in trying to distil computationally useful consequences from the strong but somewhat mysterious assumption that a given comparative tableau is OT-compatible. The proof of the implication from L-compatibility to OT-compatibility is basically identical to the proof of claim 19 in subsection 5.4.1; the proof of the reverse implication from OT-compatibility to L-compatibility is basically identical to the proof of claim 20 in subsection 5.4.2. A couple of remarks at the end of the section spell out the intuition behind the proof. In the next section, I will discuss an obvious but important consequence of this claim 41.

Claim 41 Given a comparative tableau \mathbf{A} , consider the numerical matrix $\bar{\mathbf{A}}$ defined row by row as in (409): given a row \mathbf{a} of \mathbf{A} , we construct the corresponding row $\bar{\mathbf{a}}$ of $\bar{\mathbf{A}}$ by replacing every entry equal to W by a 1, every entry equal to E by a 0 and every entry equal to L by $-w(\mathbf{a})$, where $w(\mathbf{a})$ is the number of entries equal to W in the row \mathbf{a} .

$$(409) \quad \mathbf{a} = [a_1 \quad \dots \quad a_k \quad \dots \quad a_n] \implies \bar{\mathbf{a}} = [\bar{a}_1 \quad \dots \quad \bar{a}_k \quad \dots \quad \bar{a}_n]$$

$$a_k \in \{L, E, W\} \quad \bar{a}_k \doteq \begin{cases} +1 & \text{if } a_k = W \\ 0 & \text{if } a_k = E \\ -w(\mathbf{a}) & \text{if } a_k = L \end{cases}$$

The Ranking problem (410) is equivalent to the new problem (411), in the sense that the set of rankings that solve (410) coincides with the set of refinements of ranking vectors that solve (411). If the given comparative tableau \mathbf{A} has a unique L per row (recall that by claim 2 I can transform a given comparative tableau \mathbf{A} into another tableau \mathbf{A}' OT-equivalent to \mathbf{A} such that \mathbf{A}' has a unique entry equal to L per row), then the condition that θ be nonnegative can be dropped in the statement of problem (411) while retaining the equivalence with problem (410).

$$(410) \quad \begin{array}{l} \text{given:} \quad \text{an OT-compatible comparative tableau } \mathbf{A}; \\ \text{find:} \quad \text{a ranking } \gg \text{ that is OT-compatible with } \mathbf{A}. \end{array}$$

- (411) *given:* an OT-compatible comparative tableau A ;
find: a nonnegative ranking vector θ that is L-compatible with the corresponding derived matrix \bar{A} defined in (409).

More in general, given a comparative row $\mathbf{a} = (a_1, \dots, a_k, \dots, a_n) \in \{L, E, W\}^n$, let me say that a numerical vector $\bar{\mathbf{a}} = (\bar{a}_1, \dots, \bar{a}_k, \dots, \bar{a}_n) \in \mathbb{R}^n$ is DERIVED from the comparative row \mathbf{a} iff condition (412) holds, namely every entry equal to W in \mathbf{a} corresponds to a positive entry in $\bar{\mathbf{a}}$, every entry equal to E in \mathbf{a} corresponds to a zero entry in $\bar{\mathbf{a}}$, and every entry equal to L in \mathbf{a} corresponds to a null or negative entry in $\bar{\mathbf{a}}$. Given a comparative tableau $A \in \{L, E, W\}^{m \times n}$ with m rows and n columns, let me say that a numerical matrix $\bar{A} \in \mathbb{R}^{m \times n}$ with the same number of rows and columns is DERIVED from the comparative tableau A iff each row of \bar{A} is derived from the corresponding row of A according to definition (412).

- (412) For $k = 1, \dots, n$:
- if $a_k = W$, then $\bar{a}_k > 0$;
 - if $a_k = E$, then $\bar{a}_k = 0$;
 - if $a_k = L$, then $\bar{a}_k \leq 0$.

Then, a comparative tableau A is OT-compatible iff every numerical matrix \bar{A} derived from A is L-compatible with a ranking vector with nonnegative components.

First part of the proof. Consider a nonnegative ranking vector θ that solves the problem (411), namely such that $\langle \theta, \bar{\mathbf{a}} \rangle > 0$ for every row $\bar{\mathbf{a}}$ of the numerical matrix \bar{A} defined in (409). Let me show that θ is OT-compatible with the comparative tableau A , so that every refinement of θ solves problem (410). The following chain of inequalities (413) holds for every row \mathbf{a} of the comparative tableau A and every $k \in L(\mathbf{a})$. This chain of inequalities is basically identical to the one in (163). Here, I have reasoned as follows: in step (a), I have used the hypothesis that θ solves problem (411); in step (b), I have used the definition (94) of the scalar product $\langle \theta, \bar{\mathbf{a}} \rangle$; in step (c), I have split up the set $\{1, \dots, n\}$ that h runs over into the three sets $W(\mathbf{a})$, $L(\mathbf{a})$ and their complement; in step (d), I have noted that $\bar{a}_h = 1$ for every $h \in W(\mathbf{a})$, that $\bar{a}_h = -w(\mathbf{a})$ for every $h \in L(\mathbf{a})$ and that $\bar{a}_h = 0$ for every $h \notin W(\mathbf{a}) \cup L(\mathbf{a})$, by the definition (409) of $\bar{\mathbf{a}} = (\bar{a}_1, \dots, \bar{a}_n)$; in step (e), I have upper bounded the sum $\sum_{h \in W(\mathbf{a})} \theta_h$ with its biggest term $\max_{h \in W(\mathbf{a})} \theta_h$ multiplied by the number $w(\mathbf{a})$ of terms; in step (f), I have used the hypothesis that all the components of θ are nonnegative and thus $\sum_{h \in L(\mathbf{a})} \theta_h \geq \theta_k$ provided that $k \in L(\mathbf{a})$.

$$\begin{aligned}
 (413) \quad 0 & \stackrel{(a)}{<} \langle \theta, \bar{\mathbf{a}} \rangle \\
 & \stackrel{(b)}{=} \sum_{h=1}^n \theta_h \bar{a}_h \\
 & \stackrel{(c)}{=} \sum_{h \in W(\mathbf{a})} \theta_h \bar{a}_h + \sum_{h \in L(\mathbf{a})} \theta_h \bar{a}_h + \sum_{h \notin W(\mathbf{a}) \cup L(\mathbf{a})} \theta_h \bar{a}_h \\
 & \stackrel{(d)}{=} \sum_{h \in W(\mathbf{a})} \theta_h - w(\mathbf{a}) \sum_{h \in L(\mathbf{a})} \theta_h + 0 \\
 & \stackrel{(e)}{\leq} w(\mathbf{a}) \max_{h \in W(\mathbf{a})} \theta_h - w(\mathbf{a}) \sum_{h \in L(\mathbf{a})} \theta_h \\
 & \stackrel{(f)}{\leq} w(\mathbf{a}) \max_{h \in W(\mathbf{a})} \theta_h - w(\mathbf{a}) \theta_k
 \end{aligned}$$

By reordering the inequality (413), I obtain $\max_{h \in W(\mathbf{a})} \theta_h > \theta_k$. Since this conclusion holds for every $k \in L(\mathbf{a})$, then I also have $\max_{h \in W(\mathbf{a})} \theta_h > \max_{k \in L(\mathbf{a})} \theta_k$. This conclusion says that θ satisfies condition (405a) and is thus OT-compatible with the comparative row \mathbf{a} . Since this conclusion holds for every row \mathbf{a} , then θ is OT-compatible with the comparative tableau A . If the row \mathbf{a} has a unique entry equal to L , then we can drop the assumption that the ranking vector θ be nonnegative

The chain of implications in (416) says that θ satisfies condition (408) and is thus L-compatible with \bar{a} . Since this conclusion holds for every row \bar{a} , then θ is L-compatible with the numerical matrix \bar{A} . Note that this reasoning does not depend on the specific definition (409) of the numerical matrix \bar{A} , but only on the fact that the entries in \bar{A} corresponding to w's in A are positive. Thus, this reasoning extends from the specific derived matrix (409) to an arbitrary derived matrix (412), thus showing that all derived matrices of an OT-compatible comparative tableau are L-compatible. ■

Why is the numerical matrix \bar{A} in (409) defined the way it is? In other words, what is the intuition behind pairing up entries corresponding to L's with the opposite of the number $w(a)$ of winner-prefering constraints? In order to bring out the intuition, let's restrict ourselves to the case of comparative tableaux that have a unique entry equal to L per row. To illustrate, I give in (417) the derived matrix (409) corresponding to Pater's tableau (82). For the sake of readability, I am omitting E's in A and O's in \bar{A} . Note that the numerical matrix \bar{A} in (417) has the following crucial property: for every row $\bar{a} = (\bar{a}_1, \dots, \bar{a}_5)$, the sum $\sum_{i=1}^5 \bar{a}_i$ over its corresponding entries is null. This is a property that holds of the numerical matrix \bar{A} defined in (409) in the general case, and it is ensured by the fact that the L of a comparative row a is replaced precisely by $-w(a)$. Let me now point out the intuition for why this property that rows sum up to zero is crucial.

$$(417) \quad A = \begin{bmatrix} W & L & W & & \\ & W & L & W & \\ & & W & L & W \\ & & & W & L \end{bmatrix} \implies \bar{A} = \begin{bmatrix} +1 & -2 & +1 & & \\ & +1 & -2 & +1 & \\ & & +1 & -2 & +1 \\ & & & +1 & -1 \end{bmatrix}$$

Let e be the vector of \mathbb{R}^n with n components all identical to 1. Thus, $\theta' = \theta + \lambda e$ denotes the ranking vector obtained from the ranking vector θ by summing the same constant $\lambda \in \mathbb{R}$ to each one of its components. Of course, if a vector θ is OT-compatible with a comparative tableau A , then the vector $\theta' = \theta + \lambda e$ is OT-compatible with A too for every $\lambda \in \mathbb{R}$, since OT-compatibility is not sensitive to the absolute size of the components of a ranking vector, but only to their relative sizes. Thus, in order for L-compatibility with some derived numerical matrix \bar{A} to entail OT-compatibility with the original comparative tableau A , it is necessary to define \bar{A} in such a way that, if a vector θ is L-compatible with \bar{A} , then the vector $\theta' = \theta + \lambda e$ is L-compatible with \bar{A} too for every $\lambda \in \mathbb{R}$. The numerical matrix \bar{A} satisfies this property, as shown in (418): the crucial property that the rows of \bar{A} sum up to zero ensures that step (*) holds.

$$(418) \quad \begin{aligned} \langle \theta + \lambda e, \bar{a} \rangle &= \langle \theta, \bar{a} \rangle + \langle \lambda e, \bar{a} \rangle \\ &= \langle \theta, \bar{a} \rangle + \lambda \underbrace{\sum_{k=1}^n \bar{a}_k}_{\text{null}} \\ &\stackrel{(*)}{=} \langle \theta, \bar{a} \rangle \end{aligned}$$

Here is another way of making sense of the definition (409) of the derived matrix \bar{A} used in the first part of the preceding proof. Let a LOSS FUNCTION Loss be a function that takes a piece of data and a ranking vector θ and returns 0 if θ accounts for that piece of data and 1 otherwise. In the case of standard and linear OT, the piece of data are respectively a comparative row a and a numerical vector \bar{a} , and the conditions for succes are respectively OT-compatibility and L-compatibility. Thus, we get the definitions in (419).

$$(419) \quad \begin{aligned} \text{a. } \text{Loss}_L(\bar{a}, \theta) &= \begin{cases} 1 & \text{if } \langle \theta, \bar{a} \rangle \leq 0 \\ 0 & \text{otherwise} \end{cases} \\ \text{b. } \text{Loss}_{OT}(\bar{a}, \theta) &= \begin{cases} 1 & \text{if } \max_{h \in W(a)} \theta_h - \max_{k \in L(a)} \theta_k \leq 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

The idea of the derived numerical matrix \bar{A} is to ensure that the inequality in (420a) holds for every comparative row a and the corresponding derived numerical vector \bar{a} . This inequality in turn trivially entails the inequality (420b.ii).

$$(420) \quad \text{a. } \langle \theta, \bar{\mathbf{a}} \rangle \leq w(\mathbf{a}) \cdot \left(\max_{h \in W(\mathbf{a})} \theta_h - \max_{k \in L(\mathbf{a})} \theta_k \leq 0 \right)$$

$$\text{b. } 0 \stackrel{(i)}{\leq} \text{Loss}_{\text{OT}}(\bar{\mathbf{a}}, \theta) \stackrel{(ii)}{\leq} \text{Loss}_L(\bar{\mathbf{a}}, \theta)$$

The inequality (420b) says that the problem of minimizing the OT-loss can be replaced by the problem of minimizing the L-loss.

8.2 Computational consequences

Because of the fact that the core notion of OT-compatibility (405) is stated in terms of the maximum operator, standard OT displays the crucial property of *strict domination*, according to which the highest ranked relevant constraint “takes it all”. Because of this property, OT looks *prima facie* like a rather exotic combinatorial framework. Exotic in the sense that it does not seem to have any close correspondent within mainstream Learning Theory. For this reason, computational OT has been developed in the current literature along the lines described in (421). The literature reviewed in section 4.2.1 exemplifies well the classical approach (421) to computational OT.

(421) *The classical approach to computational OT.* Computational problems that arise in modeling the acquisition of phonology within the framework of OT are tackled by means of *ad hoc* combinatorial algorithms tailored to the exotic framework of OT, developed from scratch with no connections to methods and results from mainstream Learning Theory.

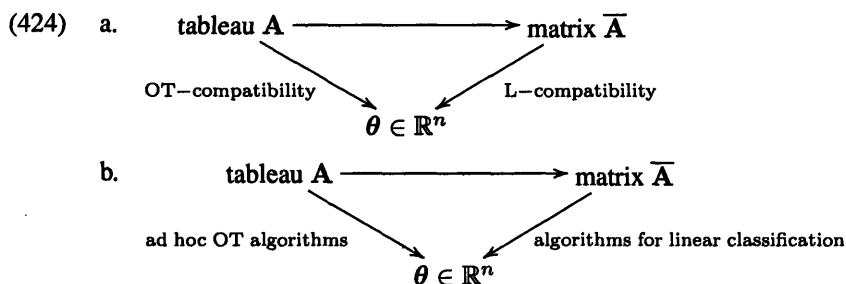
In order to bridge this gap between computational OT and mainstream Learning Theory, various scholars have recently started to explore the hypothesis of replacing standard OT by linear OT, since the latter falls within the general class of *linear models* very well studied in mainstream Learning Theory. Some recent examples of this line of research are Hayes and Wilson (2008), Coetzee and Pater (2008), Boersma and Pater (2007, 2008), Pater et al. (2006), Jesney and Tessier (2007, 2008), Jäger (2007), Potts et al. (2008), Goldwater and Johnson (2003), etc.

(422) Linear OT is superior to standard OT because it comes with well established algorithms from the theory of linear classification, that standard OT does not.

Here are some quotes that document (422), from Pater (2009): “[I will] illustrate and extend existing arguments for the replacement of [standard] OT’s ranked constraints with [linear OT’s] weighted ones: that the resulting theory can be adapted relatively straightforwardly to deal with various types of non-categorical linguistic phenomena, and that it can make use of well understood algorithms for the modeling of learning and for other computational implementations. [...] The strengths of [linear OT over standard OT] in this area are of considerable importance” (p.3); “[linear OT] allows the use of existing, well-understood algorithms [such as the Simplex Method] for computational implementations” (p. 9); “One broad argument for [linear OT] is that [...] [it is] compatible with existing well-understood algorithms for learning variable outcomes and for learning gradually [...]”. Since these algorithms are broadly applied with connectionist and statistical models of cognition, this forms an important connection between the [linear OT] version of generative linguistics and other research in cognitive science” (pp. 18-19). In this section, I want to point out that claim (422) is *wrong*, as stated in (423).

(423) Linear OT is *not* in any way computationally superior to standard OT, since pretty much any algorithm for linear OT can be readapted to standard OT through the equivalence between OT-compatibility and L-compatibility guaranteed by claim 41.

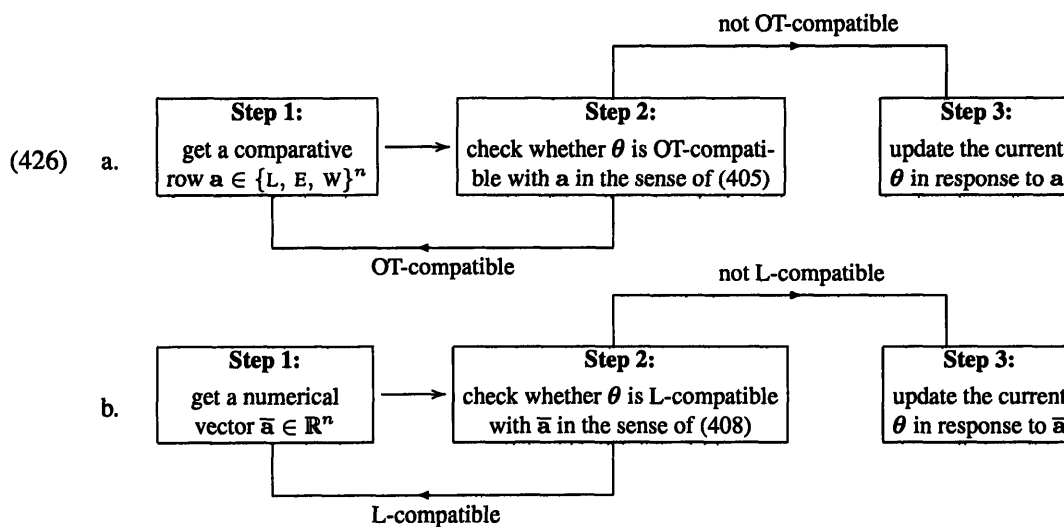
Let me elaborate on claim (423). Given a comparative tableau \mathbf{A} , I have paired it up with the numerical matrix $\bar{\mathbf{A}}$ in (409). By working at the level of the comparative tableau \mathbf{A} , we have the notion of OT-compatibility in (405). By working at the level of the numerical matrix $\bar{\mathbf{A}}$, we have the notion of L-compatibility in (408). We are thus confronted with the situation in (424a). The preceding claim 41 ensures that it does not make a difference at which level we decide to work.



Thus, there are two strategies to devise algorithms for computational OT, as described in (424b). One strategy is to work at the level of comparative tableaux, and devise *ad hoc* combinatoric algorithms for OT-compatibility. This strategy yields the classical approach described in (421), which has been explored in the literature as illustrated in section 4.2.1. Another strategy is to rephrase problems in computational OT in terms of L-compatibility and import within OT algorithms for linear classification. The latter strategy yields the alternative approach to computational OT described in (425).

(425) *An alternative approach to computational OT.* Rather than devising from scratch *ad hoc* combinatorial algorithms, computational problems that arise in modeling the acquisition of phonology within the framework of OT can be tackled by importing and straightforwardly adapting well known algorithms from the theory of linear classifiers developed within mainstream Learning Theory.

I think that the alternative approach (425) might have significant implications. Section 5.4 has provided a first application of this alternative approach (425). Let me make this point explicit. The general shape of the on-line algorithm for standard OT is repeated once more in (426a). In step 1, the algorithm receives a comparative row $\mathbf{a} \in \{L, E, W\}^n$; in step 2, the algorithm checks whether the current ranking vector θ is OT-compatible with that comparative row \mathbf{a} , namely whether condition (405) holds; if it doesn't, then the algorithm takes action in step 3. Analogously, the general shape of the on-line algorithm for linear OT is repeated once more in (426b). In step 1, the algorithm receives an n -tuple $\bar{\mathbf{a}} = (\bar{a}_1, \dots, \bar{a}_k, \dots, \bar{a}_n) \in \mathbb{R}^n$ of numbers; in step 2, the algorithm checks whether the current ranking vector θ is L-compatible with the numerical vector $\bar{\mathbf{a}}$, namely whether condition (408) holds; if it doesn't, then the algorithm takes action in step 3.

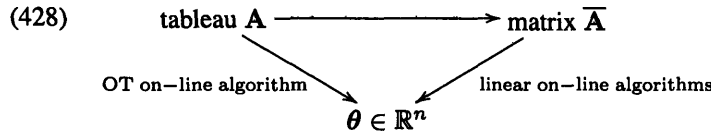


Algorithms of the form (426b) are very well studied in the field of linear classification; see for instance Cesa-Bianchi and Lugosi (2006, Chp. 12) for a modern introduction. The parallelism between the standard OT on-line algorithm (426a) and the linear OT on-line algorithm (426b) suggests the following strategy to obtain convergent update rules for step 3 of the standard OT on-line algorithm (426a). Assume that the linear OT on-line algorithm (426b) converges with a given update rule in

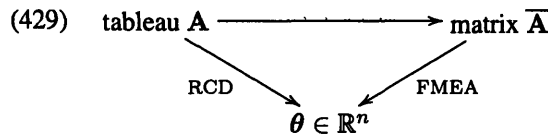
step 3 under certain hypotheses on the input set of numerical vectors that the data in step 1. If it were possible to devise a mapping $a \in \{L, E, W\}^n \mapsto \bar{a} \in \mathbb{R}^n$ from comparative rows into numerical vectors that satisfies the two properties (427), then the standard OT on-line algorithm (426a) with that same update rule in step 3 converges too. Claim 41 ensures that such a mapping exists.

- (427) a. If a comparative row a is not OT-compatible with a ranking vector θ , then the corresponding numerical vector \bar{a} is not L-compatible with that same ranking vector θ .
- b. If a set of comparative rows is OT-compatible, then the set of corresponding numerical vectors satisfy all the hypotheses required for the convergence of the linear OT on-line algorithm (426b) with the update rule considered.

In other words, provably convergent promotion-demotion update rules for the OT on-line algorithm can be obtained by readapting update rules for the linear on-line algorithm, as schematized in (428).



To illustrate another application of the approach (425), let me point out that Tesar’s RCD (373) for the Ranking problem in OT can be reinterpreted as the Fourier Motzkin Elimination Algorithm (henceforth: FMEA) for polyhedral feasibility, as depicted in (429). Let me sketch the details.



A polyhedron is a set of ranking vectors $\theta \in \mathbb{R}^n$ that satisfy a finite number of linear inequalities, as in (430a). A polyhedral feasibility problem is the following problem: given (430a), establish whether P is empty or not; in the latter case, find a ranking vector θ that belongs to P . Here is a natural strategy to solve polyhedral feasibility problems. We know that a set is nonempty iff its projection along any coordinate is nonempty. Furthermore, every time we project down a set, we reduce the dimension of the problem by 1 and thus potentially simplify the problem. Thus, to establish whether a polyhedron is empty or not, one might keep projecting it down until dimension 1 is reached, and then solve the feasibility problem for the 1-dimensional problem thus obtained. This is precisely the idea of the FMEA. Yet, in the general case, this is not a very efficient strategy to solve polyhedral feasibility problems. Here is the reason. Given an arbitrary polyhedron $P \subseteq \mathbb{R}^n$ as in (430a), its orthogonal projection $P' \subseteq \mathbb{R}^{n-1}$ along, say, the n th axis is given in (430b).¹ Thus, each time we project down and thus reduce the dimension of the problem by 1, we run into the danger of adding lots and lots of new linear inequalities, namely all the inequalities of type II in (430).

$$(430) \quad a. \quad P = \left\{ \theta = (\theta_1, \dots, \theta_n) \in \mathbb{R}^n \mid \langle \theta, \bar{a}_i \rangle = \sum_{j=1}^n \bar{a}_{i,j} \theta_j \geq 1, \quad i = 1, \dots, m \right\}$$

inequalities of type I

$$b. \quad P' = \left\{ (\theta_1, \dots, \theta_{n-1}) \mid \left. \begin{array}{l} \sum_{i=1}^{n-1} \bar{a}_{k,i} \theta_i \geq 1 \quad \text{for } k \text{ s.t. } \bar{a}_{k,n} = 0 \\ \sum_{i=1}^{n-1} \left(\frac{\bar{a}_{k,i}}{\bar{a}_{k,n}} - \frac{\bar{a}_{h,i}}{\bar{a}_{h,n}} \right) \theta_i \geq \frac{1}{\bar{a}_{k,n}} - \frac{1}{\bar{a}_{h,n}} \quad \text{for } k, h \text{ s.t. } \bar{a}_{k,n} > 0 \\ \text{and } \bar{a}_{h,n} < 0 \end{array} \right\}$$

inequalities of type II

¹This is a straightforward computation; see for instance Bertsimas and Tsitsiklis (1997, pp. 70-74).

Given an instance (410) of the Ranking problem, claim 41, ensures that we can equivalently solve the corresponding problem (411). The latter problem is of course a polyhedral feasibility problem. So we can tackle it by the FMEA. Crucially, no inequalities of type II ever arise when the polyhedron (430a) is defined through the numerical matrix $\bar{\mathbf{A}}$ defined in (409) corresponding to an OT-compatible comparative tableau. In fact, if the comparative tableau \mathbf{A} is OT-compatible, then the corresponding matrix $\bar{\mathbf{A}}$ has at least one column, say the n th column, whose entries are all nonnegative, namely such that we cannot find two entries $\bar{a}_{h,n}$ and $\bar{a}_{k,n}$ with different sign. Thus, if we project along that dimension, then we reduce the dimension of the problem by 1 without increasing the number of linear constraints, because we get no linear constraints of type II. Instead, we get only inequalities of type I that indeed correspond to the simplified tableau considered by RCD.

8.3 More consequences

The parallelism between the two frameworks of standard and linear OT reviewed at the beginning of the chapter can be brought out in a particularly vivid way under the auxiliary hypothesis that all the constraints C_1, \dots, C_n are binary, namely they can only take the values 0 and 1, so that the corresponding constraint differences can only take values $-1, 0$ or 1 . Under this hypothesis, we have that a ranking vector $\theta = (\theta_1, \dots, \theta_n)$ is L-compatible with an underlying/winner/loser form triplet iff condition (431) holds for the corresponding comparative row \mathbf{a} .

$$(431) \text{ Linear OT: } \sum_{h \in W(\mathbf{a})} \theta_h > \sum_{k \in L(\mathbf{a})} \theta_k.$$

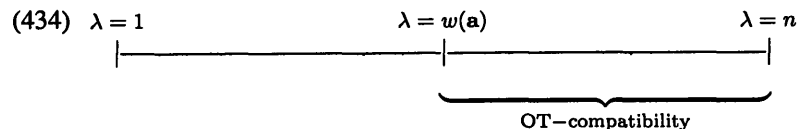
Condition (431) is completely parallel to the condition (432) for OT-compatibility between a ranking vector $\theta = (\theta_1, \dots, \theta_n)$ and a comparative row \mathbf{a} , only with the maximum operator replaced by the summation operator.

$$(432) \text{ Standard OT: } \max_{k \in W(\mathbf{a})} \theta_k > \max_{h \in L(\mathbf{a})} \theta_h.$$

Let's restrict ourselves to nonnegative ranking vectors $\theta = (\theta_1, \dots, \theta_n) \geq 0$ (or, alternatively, to comparative tableaux that have a unique entry equal to L per row). Consider now the new condition (433), dependent on the positive parameter $\lambda > 0$. Of course, if we pick λ "small", namely $\lambda = 1$,² then condition (433) boils down to the condition (431) of L-compatibility. Claim 41 says that if we pick λ "large" enough, then condition (433) entails the condition (432) that goes into the notion of OT-compatibility. As a starting point, I have considered in (409) the smallest value of λ that ensures that (433) entails (432), namely the number $\lambda = w(\mathbf{a})$ of W's in the row \mathbf{a} ; of course, any value larger than that is just as good; thus a suitable candidate is $\lambda = n$.

$$(433) \text{ A restatement of standard OT: } \sum_{k \in W(\mathbf{a})} \theta_k > \lambda \sum_{h \in L(\mathbf{a})} \theta_h$$

For any $\lambda \geq 1$, let me say that a ranking vector θ is λ -COMPATIBLE with a comparative row \mathbf{a} iff the condition (433) holds for the corresponding value of λ . We can slide from linear OT to standard OT just by increasing the parameter λ in the condition (433) from $\lambda = 1$ to any value larger than $w(\mathbf{a})$.



If the correct model of phonology were linear OT, then phonology would fit squarely within mainstream Learning Theory: the constraints map the data into the inner product space \mathbb{R}^n and any phonology is a linear model within such space. Suppose instead that the right model of phonology is standard OT. Why should this be? I want to speculate that standard OT has indeed *two* advantages over linear OT. The *first* advantage is that standard OT posits a model of constraint interaction very

²The choice $\lambda < 1$ does not make any sense.

simple, that thus allows restrictiveness to be approximately described by compact measures such as Prince and Tesar's measure (352). I defer a discussion of this first advantage to future work. The *second* advantage concerns processing. Every time the speaker wants to utter an underlying form x , he needs to compute the corresponding surface form according to his own phonology. Suppose that the right model of phonology were linear OT. Then in production the speaker would have to compute many scalar products $\langle \theta, \bar{a} \rangle$ between the vector θ that represents his grammar and the vectors \bar{a} of constraints violations corresponding to various underlying/winner/loser form triplets. The computation of one such scalar product $\langle \theta, \bar{a} \rangle$ requires only linear time in the number n of constraints. Yet, the number of constraints n needed to account for cross-linguistic variation might plausibly be very large. Furthermore, it might also be plausible to assume that the vectors of constraint differences \bar{a} are not sparse, namely they do not contain many null entries. Thus, it might be too costly for the speaker to compute all these scalar products exactly. In the face of this problem, the speaker might want to settle for an approximation of the values of these scalar products. Here is a sensible way to approximate one of these scalar products $\langle \theta, \bar{a} \rangle$. First, we can of course throw away all components θ_k of the ranking vector that correspond to null constraint differences $\bar{a}_k = 0$ as in (435a), because the product $\theta_k \bar{a}_k$ is null and thus does not contribute to the overall scalar product. Next, we could split up the remaining components of the ranking vector as in (435b) into the set of large components and the set of small components; finally, we could approximate the scalar product $\langle \theta, \bar{a} \rangle$ as in (435c), by throwing away the products $\theta_k \bar{a}_k$ corresponding to small components θ_k of the ranking vector and using just what is left.

$$\begin{aligned}
 (435) \quad \langle \theta, \bar{a} \rangle &= \sum_{k=1}^n \theta_k \bar{a}_k \\
 &\stackrel{(a)}{=} \sum_{k=1, \bar{a}_k \neq 0}^n \theta_k \bar{a}_k \\
 &\stackrel{(b)}{=} \sum_{k=1, \bar{a}_k \neq 0, \theta_k \text{ large}}^n \theta_k \bar{a}_k + \sum_{k=1, \bar{a}_k \neq 0, \theta_k \text{ small}}^n \theta_k \bar{a}_k \\
 &\stackrel{(c)}{\approx} \sum_{k=1, \bar{a}_k \neq 0, \theta_k \text{ large}}^n \theta_k \bar{a}_k
 \end{aligned}$$

Standard OT arises from linear OT this way, when we make the drastic choice of retaining only the largest component θ_k corresponding to nonnull constraint differences.³

³See Eisner (2000, par. 2) makes a somewhat related point, as pointed out to me by Adam Albright (p.c.).

List of symbols used in part II

\mathbb{R}	set of real numbers
\mathcal{X}	set of underlying forms
x	generic underlying form
\mathcal{Y}	set of candidate surface forms
y^*	intended winner surface form
y	intended loser surface form
Gen	generating function
\mathcal{C}	constraint set
C_k	generic constraint, namely a function from $\mathcal{X} \times \mathcal{Y}$ into natural numbers
n	total number of constraints
\gg	ranking of the constraint set
OT_{\gg}	OT-grammar corresponding to the ranking \gg
$\mathcal{R}(OT_{\gg})$	range of the OT-grammar OT_{\gg} , namely the language corresponding to the ranking \gg
\mathbf{A}	comparative tableau
\mathbf{a}	comparative row, namely an arbitrary row of a comparative tableau
$dec_{\gg}(\mathbf{a})$	decisive constraint for the comparative row \mathbf{a} w.r.t. the ranking \gg
$dec_{\gg}(C_k)$	set of comparative rows whose decisive constraint w.r.t. \gg is C_k
$W(\mathbf{a})$	set of constraints that have a W in the comparative row \mathbf{a}
$L(\mathbf{a})$	set of constraints that have an L in the comparative row \mathbf{a}
$w(\mathbf{a})$	number of W's in the comparative row \mathbf{a}
θ	ranking vector, namely an n -tuple of numbers
θ_k	ranking value of constraint C_k , namely the k th component of the ranking vector θ
$L(\mathbf{a}, \theta)$	set of loser-preferring constraints w.r.t. \mathbf{a} ranked by θ above the highest ranked winner-preferring constraint
$\ell(\mathbf{a}, \theta)$	cardinality of the set $L(\mathbf{a}, \theta)$
$\max_{k \in \Omega} \theta_k$	largest ranking value θ_k among those that correspond to constraints C_k with $k \in \Omega$
$\sum_{k \in \Omega} \theta_k$	sum of the ranking values θ_k that correspond to constraints C_k with $k \in \Omega$
$\bar{\mathbf{a}}$	numerical vector corresponding to the comparative row \mathbf{a}
$\{\mathbf{L}, \mathbf{E}, \mathbf{W}\}^n$	set of comparative rows (for n constraints)
$\{\mathbf{L}, \mathbf{E}, \mathbf{W}\}^{m \times n}$	set of comparative tableaux (with m rows and n columns)
$\langle \cdot, \cdot \rangle$	scalar product between two vectors

Bibliography

- Abrusan, Marta. 2009. "Presuppositional and Negative Islands: A Semantic Account". Institut Jean Nicod.
- Abrusan, Marta, and Benjamin Spector. 2009. "An Interval Based Semantics for Negative Degree Questions". Institut Jean Nicod.
- Albright, Adam, Giorgio Magri, and Jennifer Michael. 2007. "Modeling Doubly Marked Lags with Additive Constraint Interaction". In *Proceedings of BUCLD 2007*, -. .
- Amsili, Pascal, and Claire Beyssade. 2006. "Compulsory Presuppositions in Discourse". In *Proceedings of the Workshop on Constraints in Discourse*, -. .
- Barwise, J., and R. Cooper. 1991. "Generalized Quantifiers and Natural Language". *Linguistics and Philosophy* 4.2:159–220.
- Bennett, Michael, and Barbara Partee. 1972. "Toward the Logic of Tense and Aspect in English". Distributed in 1978 by the Indiana University Linguistics Club, Bloomington.
- Berman, Steve. 1990. "Toward the Semantics of Open Sentences: *Wh*-Phrases and Indefinites". In *Proceedings of the Seventh Amsterdam Colloquium*, ed. Martin Stokhof and Leen Torenvliet, 53–78. University of Amsterdam: ITLI.
- Bernhardt, Barbara Handford, and Joseph Paul Stemberger. 1998. *Handbook of phonological development from the perspective of constraint-based nonlinear phonology*. Academic Press.
- Bertsimas, Dimitris, and John N. Tsitsiklis. 1997. *Linear Optimization*. Athena Scietific.
- Berwick, Robert. 1985. *The acquisition of syntactic knowledge*. Cambridge, MA: MIT Press.
- Boersma, Paul. 1997. "How We Learn Variation, Optionality and Probability". In *IFA Proceedings 21*, 43–58. University of Amsterdam: Institute for Phonetic Sciences.
- Boersma, Paul. 1998. Functional Phonology. Doctoral Dissertation, University of Amsterdam. The Hague: Holland Academic Graphics.
- Boersma, Paul. 1999. "Optimality Theoretic Learning in the Praat Program". In *IFA Proceedings 23*, 17–35. University of Amsterdam: Institute for Phonetic Sciences. .
- Boersma, Paul. 2008. "Some Correct Error-driven Versions of the Constraint Demotion Algorithm". To appear in *Linguistic Inquiry*.
- Boersma, Paul, and Bruce Hayes. 2001. "Empirical Tests for the Gradual Learning Algorithm". *Linguistic Inquiry* 32:45–86.
- Boersma, Paul, and Clara Levelt. 2000. "Gradual Constraint-Ranking Learning Algorithm Predicts Acquisition Order". In *Proceedings of the 30th Child Language Research Forum*, 229–237. Standford University: CSLI. Corrected version (ROA 361, 1999/08/28).

- Boersma, Paul, and Joe Pater. 2007. "Convergence Properties of a Gradual Learner for Harmonic Grammar". In *Proceedings of NELS 38*, -. .
- Boersma, Paul, and Joe Pater. 2008. "Convergence Properties of a Gradual Learning Algorithm for Harmonic Grammar". Ms.
- Burton, S., and J. Grimshaw. 1992. "Coordination and VP-internal subjects". *Linguistic Inquiry* 23:305–313.
- Carlson, Gregory N. 1977. Reference to Kinds in English. Doctoral Dissertation, University of Massachusetts at Amherst. Published in 1980 by Garland Press, New York.
- Cesa-Bianchi, Nicoló. 1997. "Analysis of two Gradient-based Algorithms for On-line Regression". In *Proceedings of COLT 97*, -. .
- Cesa-Bianchi, Nicoló, and Gábor Lugosi. 2006. *Prediction, Learning, and Games*. Cambridge University Press.
- Chemla, Emmanuel. 2006. "A Problem for the Theory of Anti-Presuppositions". Manuscript, ENS Paris.
- Chemla, Emmanuel. 2008. "An Epistemic Step for Anti-Presuppositions". *Journal of Semantics* 25:141–173.
- Chierchia, Gennaro. 1988. Topics in the Syntax and Semantics of Infinitives and Gerunds. Doctoral Dissertation, University of Massachusetts, Amherst.
- Chierchia, Gennaro. 1994. "A Note on the Individual-Level vs. Stage-Level Contrast in German". In *Proceedings of the Cortona Meeting on Tense and Aspect*, 111–123. Turin: Rosenberg and Sellier.
- Chierchia, Gennaro. 1995. "Individual-level Predicates as Inherent generics". In *The Generic Book*, ed. G. N. Carlson and F. J. Pelletier, 125–175. The Univ. of Chicago Press.
- Chierchia, Gennaro. 1998. "Reference to Kinds across Languages". *Natural Language Semantics* 6:339–405.
- Chierchia, Gennaro. 2004. "Scalar implicatures, polarity phenomena and the syntax/pragmatics interface". In *Structures and Beyond*, ed. A. Belletti. Oxford: Oxford University Press.
- Chierchia, Gennaro. 2006a. "Broaden Your Views: Implicatures of Domain Widening and the 'Logicity' of Language". *Linguistic Inquiry* 37.4:535–590.
- Chierchia, Gennaro. 2006b. "Obligatory Implicatures". Talk delivered at *Sinn und Bedeutung 11*, Universitat Pompeu Fabra, Barcelona.
- Chierchia, Gennaro. 2007. "Polarity Sensitivity: The Role of Higher-Order Implicatures". In *Proceedings of SALT 17*, -. .
- Chierchia, Gennaro, Danny Fox, and Benjamin Spector. 2007. "Hurtford's Constraints and the Theory of Scalar Implicatures". Talk delivered at the "Paris-MIT Workshop on Presuppositions and Implicatures".
- Chierchia, Gennaro, Danny Fox, and Benjamin Spector. to appear. "The Grammatical View of Scalar Implicatures and the Relationship between Semantics and Pragmatics". In *Handbook of Semantics*, ed. Paul Portner, Claudia Maienborn, and Klaus von Heusinger. Berlin, New York: Mouton de Gruyter.
- Coetzee, Andries W., and Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in muna and arabic. *Natural Language and Linguistic Theory* 26:289–337.

- Cohen, A. 2005. "More Than Bare Existence: An Implicature of Existential Bare Plurals". *Journal of Semantics* 22:389–400.
- Cohen, Ariel. 2000. "Bare Plurals and Suitability". In *From NP to DP: International Conference on the Syntax and Pragmatics of Noun Phrases*, -. Antwerp, Belgium. .
- Cohen, Ariel, and Nomi Erteschik-Shir. 2002. "Topic, Focus and the Interpretation of Bare Plurals". *Natural Language Semantics* 10:125–165.
- Cohen, L. J. 1971. "The logical particles of natural language". In *Pragmatics of Natural Language*, ed. Y. Bar-Hillel, 50–68. Dordrecht: Reidel.
- Cohen, W., William, Robert E. Schapire, and Yoram Singer. 1999. Learning to order things. *Journal of Artificial Intelligence Research* 10:243–270.
- Condoravdi, C. 1997. Descriptions in Contexts. Doctoral Dissertation, Yale University. Published 1997 by Garland Press, New York.
- Condoravdi, Cleo. 1992. "Individual-level predicates in conditional clauses". Talk given at the LSA Annual Meeting, Philadelphia, PA.
- Cormen, Thomas H., Charles E. Leiserson, and Ronald L. Rivest. 1990. *Introduction to Algorithms*. Cambridge, MA: MIT Press.
- Cristianini, Nello, and John Shawe-Taylor. 2000. *An Introduction to Support Vector Machines and Other Kernel-Based Methods*. Cambridge University Press.
- Curtin, Suzanne, and Kie Zuraw. 2002. "Explaining Constraint Demotion in a Developing System". In *Proceedings of the 26th annual Boston University Conference on Language Development*, ed. Anna H.-J. Do, Laura Domínguez, and Aimee Johansen, -. Cascadilla Press. .
- Davidson, Lisa, Peter W. Jusczyk, and Paul Smolensky. 2004. "The initial and final states: Theoretical implications and experimental explorations of richness of the base". In *Constraints in Phonological Acquisition*, ed. R. Kager, J. Pater, and W. Zonneveld, 158–203. Cambridge University Press.
- Deal, Amy Rose. 2008. "Events in Space". In *Proceedings of SALT18*, -. .
- Diesing, Molly. 1992. *Indefinites*. Cambridge, Mass.: MIT Press.
- Dobrovie-Sorin, Carmen. 2003. "Generic Plural Indefinites and (In)direct Binding". In *A Handbook of French Semantics*, ed. F. Corblin and de H. Swart, -. CSLI, Stanford.
- Dowty, David R. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Kluwer Academic Publisher.
- Eisner, Jason. 2000. "Easy and Hard Constraint Ranking in Optimality Theory". In *Finite-State Phonology: Proceedings of the Fifth Workshop of the ACL Special Interest Group in Computational Phonology (SIGPHON)*, ed. J. Eisner, L. Karttunen, and A. Thériault, 22–33. Luxembourg.
- Fernald, Theodore B. 2000. *Predicates and Temporal Arguments*. New York, Oxford: Oxford University Press.
- Ferreira, Marcelo. 2005. "Bare Habituals and Plural Definite Descriptions". In *Proceedings of SuB9*, ed. Emar Maier, Corien Bary, and Janneke Huitink, 102–115. www.ru.nl/ncs/sub9.
- Fikkert, Paula, and Helen De Hoop. 2009. "Language acquisition in optimality theory". *Linguistics* 47.2:311–357.
- von Fintel, Kai. 1993. "Exceptive Constructions". *Natural Language Semantics* 1.2:123–148.

- von Fintel, Kai. 1997. "Bare Plurals, Bare Conditionals, and 'Only'". *Journal of Semantics* 14:1–56.
- von Fintel, Kai. 2003. "Would you Believe it? The King of France is Back! (Presuppositions and Truth-Value Intuitions)". In *Descriptions and Beyond*, ed. Marga Reimer and Anne Bezuidenhout, 269–296. Oxford: Oxford University Press.
- von Fintel, Kai. 1994. Restrictions on Quantifier Domains. Doctoral Dissertation, UMass, Amherst.
- Fishburn, P. C. 1974. "Lexicographic Orders, Utilities and Decision Rules: A Survey". *Management Science* 20:1442–1471.
- Fodor, Janet Dean. 1970. The linguistic description of opaque contexts. Doctoral Dissertation, MIT.
- Fox, Danny. 1995. "Economy and Scope". *Natural Language Semantics* 3:283–341.
- Fox, Danny. 2000. *Economy and semantic interpretation*. MIT Press: Cambridge, Mass.
- Fox, Danny. 2004. "MIT Lecture Notes".
- Fox, Danny. 2007a. "Free Choice and the Theory of Scalar Implicatures". In *Presupposition and Implicature in Compositional Semantics*, ed. U. Sauerland and P. Stateva, 71–120. New York: Palgrave Macmillan.
- Fox, Danny. 2007b. "Too Many Alternatives: density, symmetry, and other predicaments". In *Proceedings of SALT17*. .
- Fox, Danny, and Martin Hackl. 2006. "The Universal Density of Measurement". *Linguistics and Philosophy* 29.5:537–586.
- Fox, Danny, and Roni Katzir. 2009. "Alternatives for implicature and focus". In *Presuppositions and Implicatures. Proceedings of the MIT-Paris Workshop*, ed. Paul Égré and Giorgio Magri, 101–110. Cambridge, MA: MIT Working Papers in Linguistics, volume 60.
- Fox, Danny, Jon Nissenbaum, and Uli Sauerland. 2001. "Association with Focus". Manuscript, MIT.
- Fox, Danny, and Benjamin Spector. 2008. "Economy and Embedded Exhaustification". MIT presentation on May 30, 2008.
- Gajewski, Jon. 2003. "On Analyticity in Natural Language". Manuscript, MIT.
- Gajewski, Jon. 2005. Neg-Raising: Polarity and Presupposition. Doctoral Dissertation, MIT.
- Gajewski, Jon, and Yael Sharvit. 2009. "In Defense of the Grammatical Approach to Local Implicatures". Manuscript, University of Connecticut.
- Galil, Zvi, and Nimrod Megiddo. 1977. "Cyclic Ordering is NP-complete". *Theoretical Computer Science* 5:179–182.
- Garey, Michael R., and David S. Johnson. 1979. *Computers and Intractability. A Guide to the Theory of NP-Completeness*. New York: W. H. Freeman and Company.
- Gazdar, Gerald. 1979. *Pragmatics: Implicature, Presupposition and Logical Form*. New York: Academic Press.
- Geurts, Bart. 2006. "Take 'five': The meaning and Use of a Number Word". In *Indefiniteness and Plurality*, ed. Liliane Tasmowski and Svetlana Voegelé, –. Amsterdam: John Benjamins.
- Geurts, Bart. 2009. "Scalar implicature and local pragmatics". *Mind and Language* 24:51–79.
- Geurts, Bart, and Nausicaa Pouscoulous. 2009a. "Embedded Implicatures?!?".

- Geurts, Bart, and Nausicaa Pouscoulous. 2009b. "No Scalar Inferences Under Embedding". In *Presuppositions and Implicatures. Proceedings of the MIT-Paris Workshop*, ed. Paul Égré and Giorgio Magri, 111–122. Cambridge, MA: MIT Working Papers in Linguistics, volume 60.
- Glasbey, Sheila. 1997. "I-level Predicates that Allow Existential Readings for Bare Plurals". In *Proceedings of SALT7*, ed. Aaron Lawson. Cornell University, Ithaca, NY.
- Gnanadesikan, Amalia E. 2004. "Markedness and Faithfulness Constraints in Child Phonology". In *Constraints in phonological acquisition*, ed. René Kager, Joe Pater, and Wim Zonneveld, 73–108. Cambridge: Cambridge University Press.
- Goldwater, Sharon, and Mark Johnson. 2003. "Learning OT constraint rankings using a Maximum Entropy Model". In *Proceedings of the Stockholm Workshop on Variation Within Optimality Theory*, ed. Jennifer Spenader, Anders Eriksson, and Östen Dahl, 111–120. Stockholm University.
- Grice, Paul. 1975. "Logic and Conversation". In *Syntax and Semantics 3: Speech Acts*, ed. P. Cole and J. Morgan, 41–58. New York: Academic: Academic Press.
- Gruber, J. 1965. *Studies in Lexical Relations*. Doctoral Dissertation, MIT. Published in 1965 by the Indiana University Linguistics Club, Bloomington.
- Hale, Mark, and Charles Reiss. 1998. "Formal and Empirical Arguments Concerning Phonological Acquisition". *Linguistic Inquiry* 29.4:656–683.
- Hawkins, John A. 1978. *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. London: Croom Helm.
- Hawkins, John A. 1991. "On (in)definite articles: implicatures and (un)grammaticality prediction". *Journal of Linguistics* 27:405–442.
- Hayes, Bruce. 2004. "Phonological Acquisition in Optimality Theory: The Early Stages". In *Constraints in Phonological Acquisition*, ed. R. Kager, J. Pater, and W. Zonneveld, 158–203. Cambridge University Press.
- Hayes, Bruce, Bruce Tesar, and Kie Zuraw. 2003. "OTSoft 2.1". Software package at <http://www.linguistics.ucla.edu/people/hayes/otsoft/>.
- Hayes, Bruce, and Colin Wilson. 2008. "A maximum entropy model of phonotactics and phonotactic learning". *Linguistic Inquiry* 39:379–440.
- Heim, Irene. 1988. *The Semantics of Definite and Indefinite Noun Phrases*. Doctoral Dissertation, University of Massachusetts, Amherst.
- Heim, Irene. 1991. "Artikel und Definitheit". In *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, ed. A. von Stechow and D. Wunderlich, 487–535. De Gruyter: Berlin.
- Heim, Irene, and Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Blackwell Textbooks in Linguistics.
- Heycock, Caroline. 1994. "Focus Projection in Japanese". In *Proceedings of NELS 24*, ed. M. Gonzalez, 157–171. GLSA Publications.
- Higginbotham, J., and G. Ramchand. 1997. "The Stage-Level/Individual-Level Distinction and the Mapping Hypothesis". In *Oxford University Working Papers In Linguistics, Philosophy and Phonetics, vol. 2*, ed. David Willis, 55–83.
- Horn, L. 1972. *On the Semantic Properties of Logical Operators in English*. Doctoral Dissertation, UCLA. Distributed by IULC.

- Horn, Laurence R. 2005. "The Border Wars: a neo-Gricean perspective". In *Where Semantics Meets Pragmatics*, ed. K. Turner and K. von Heusinger. Elsevir.
- Hurford, James R. 1974. "Exclusive or Inclusive Disjunction". *Foundations of Language* 11:409–411.
- Ippolito, Michela. 2006. "On the Meaning of 'Only'". In *Proceedings of SALT 16*, –. .
- Jackendoff, Ray S. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.
- Jäger, Gerhard. 2001. "Topic-Comment Structure and the Contrast Between Stage Level and Individual Level Predicates". *Journal of Semantics* 18:83–126.
- Jesney, Karen, and Anne-Michelle Tessier. 2007. "Re-evaluating learning biases in Harmonic Grammar". In *University of Massachusetts occasional papers 36: Papers in theoretical and computational phonology*, ed. Michael Becker.
- Jesney, Karen, and Anne-Michelle Tessier. 2008. "Gradual learning and faithfulness: consequences of ranked vs. weighted constraints". In *Proceedings of NELS38*, –.
- Johnson, K., and S. Tamioka. 1997. "Lowering and Mid-Size Clauses". In *Proc. of the 1997 Tübingen Workshop on Reconstruction*, ed. G. Katz, S. Kim, and W. Haike.
- Johnson, Kyle. 1996. "In search of the English Middle Field". UMass, Amherst.
- Johnson, Kyle. 2000. "Few dogs eat Whiskas or cats Alpo". In *UMOP 23: Issues in Semantics and Its Interface*, ed. K. Kusumoto and E. Villalta, 59–82. GLSA, University of Massachusetts, Amherst.
- Jusczyk, Peter, Paul Smolensky, and Theresa Alocco. 2002. "How English-learning infants respond to Markedness and Faithfulness constraints". *Language Acquisition* 10:31–73.
- Jäger, Gerhard. 2007. "Maximum Entropy Models and Stochastic Optimality Theory". In *Architectures, Rules, and Preferences. Variations on Themes by Joan W. Bresnan*, ed. Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling, and Chris Manning, 467–479. CSLI Publications, Stanford.
- Kadmon, N. 1987. On Unique and Non-Unique Reference and Asymmetric Quantification. Doctoral Dissertation, University of Massachusetts, Amherst.
- Kamp, J. A. W. 1981. "A Theory of Truth and Semantic Representation". In *Formal Methods in the Study of Language*, ed. J. Groenendijk, T. Janssen, and M. Stokhof, 277–321. Amsterdam: Mathematical Centre.
- Katz, Graham. 2003. "Event Arguments, Adverb Selection, and the Stative Adverb Gap". In *Modifying Adjuncts (Interface Explorations 4)*, ed. E. Lang, C. Maienborn, and C. Fabricius-Hansen, 455–474. Mouton de Gruyter.
- Katzir, Roni. 2008. "Structurally Defined Alternatives". *Linguistics and Philosophy* to appear.
- Kazazis, Kostas. 1969. "Possible evidence for (near-)underlying forms in the speech of a child". In *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society (CLS5)*, ed. Robert I. Binnick, Alice Davison, Georgia McGreen, and Jerry L. Morgan, 382–388. Chicago, IL: Chicago Linguistic Society.
- Kimball, J. 1973. "The Grammar of Existence". In *Papers from the Ninth Regional Meeting, Chicago Linguistic Society*, ed. C. Corum, 262–270. .
- Kiss, K. E. 1998. "On Generic and Existential Bare Plurals and the Classification of Predicates". In *Events and Grammar*, ed. S. Rothstein, 145–162. Dordrecht: Kluwer Academic Publisher.

- Kratzer, Angelika. 1995. "Stage-level and Individual-level Predicates". In *The Generic Book*, ed. Gregory N. Carlson and F. J. Pelletier, 125–175. The Univ. of Chicago Press.
- Kratzer, Angelika. 2007. "Situations in Natural Language Semantics". In *Stanford Encyclopedia of Philosophy*, ed. E. Zalta.
- Krifka, Manfred. 1995. "The Semantics and Pragmatics of Polarity Items". *Linguistic Analysis* 25:209–257.
- van Kuppevelt, Jan. 1996. "Inferring from Topics". *Linguistics and Philosophy* 19:393–443.
- Ladusaw, William A. 1986. "Principles of Semantic Filtering". In *Proceedings of the West Coast Conference on Formal Linguistics 5*, 129–141. .
- Landman, Fred. 1998. "Plurals and Maximization". In *Events and Grammar*, ed. S. Rothstein. Dordrecht: Kluwer.
- Landman, Fred. 2000. *Events and Plurality*. Dordrecht, Boston, London: Kluwer Academic Publishers.
- Larson, Richard, and Gabriel Segal. 1995. *Knowledge of Meaning*. MIT Press.
- Levelt, Clara C., Niels O. Schiller, and Willem J. Levelt. 2000. "The Acquisition of Syllable Types". *Language Acquisition* 8(3):237–264.
- Levelt, Clara C., and R. van de Vijver. 1998. "Syllable Types in Cross-Linguistic and Developmental Grammars". Rutgers Optimality Archive 265.
- Levinson, Stephen C. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.
- Lewis, David. 1975. "Adverbs of Quantification". In *Formal Semantics of Natural Languages*, ed. E. Keenan, 3–15. Cambridge: Cambridge University Press.
- Lin, V. 2000. "Determiner Sharing". In *Proceedings of WCCFL 19*, ed. R. Billerey and B. D. Lillehaugen, –. Somerville MA: Cascadilla Press. .
- Lin, V. 2001. "A way to undo A-movement". In *Proceedings of WCCFL 20*, ed. K. Megerdooimian and L. A. Barel, –. Somerville MA: Cascadilla Press. .
- Link, Godehard. 1983. "The Logical Analysis of Plurals and Mass Terms: A Lattice-Theoretical Approach". In *Meaning, Use and Interpretation in Language*, ed. R. Bauerle, C. Schwartze, and A. von Stechow, 302–323. Berlin: Mouton de Gruyter.
- Löbner, Sebastian. 1985. "Definites". *Journal of Semantics* 4:279–326.
- Magri, Giorgio. 2009. "The acquisition of Dutch syllable types: from linear OT to standard OT". Poster presented at the LSA Annual Meeting, San Francisco.
- Maienborn, Claudia. 2001. "On the Position and Interpretation of Locative Modifiers". *Natural Language Semantics* 9:191–240.
- Maienborn, Claudia. 2004. "A Pragmatic Explanation of the Stage Level/Individual Level Contrast in Combination with Locatives". In *Proceedings of the Western Conference of Linguistics (WECOL)*, ed. Brian Agbayani, Vida Samiian, and Benjamin Tucker, 158–170. Fresno: CSU. volume 15.
- Manzini, M. Rita, and Ken Wexler. 1987. "Parameters, Binding Theory, and Learnability". *Linguistic Inquiry* 18.3:413–444.

- McCarthy, John J. 1998. "Morpheme structure constraints and paradigm occultation". In *Proceedings of the Chicago Linguistics Society 5. Volume II: The Panels*, ed. Chaterine M. Gruber, Derrick Higgins, Kenneth Olson, and Tamra Wysocki, -. Chicago: CLS. .
- McNally, L. 1998a. "Stativity and Teliticity". In *Events and Grammar*, ed. S. Rothstein, 293–307. Dordrecht: Kluwer Academic Publisher.
- McNally, Louise. 1998b. "Existential Sentences without Existential Quantification". *Linguistics and Philosophy* 21:353–392.
- Milsark, Gary L. 1974. Existential Sentences in English. Doctoral Dissertation, MIT.
- Milsark, Gary L. 1977. "Toward an Explanation of Certain Peculiarities of the Existential ". *Linguistic Analysis* 3.1:1–29.
- Musan, Renate. 1995. On the Temporal Interpretation of Noun Phrases. Doctoral Dissertation, MIT.
- Musan, Renate. 1997. "Tense, Predicates, and Lifetime Effects". *Natural Language Semantics* 5:271–301.
- Partee, Barbara H., and M. Rooth. 1983. "Generalized Conjunction and Type Ambiguity". In *Meaning, Use and Interpretation of Language*, ed. R. Bauerle, C. Schwarze, and A. von Stechow. Berlin: de Gruyter.
- Pater, Joe. 2008. "Gradual Learning and Convergence". *Linguistic Inquiry* 39.2:334–345.
- Pater, Joe. 2009. "Weighted Constraints in Generative Linguistics". *Cognitive Science* 33:999–1035.
- Pater, Joe, Karen Jesney, and Anne-Michelle Tessier. 2006. "Phonological Acquisition as Weighted Constraint Interaction". In *Proceedings of GALANA 2*, ed. Alyona Beliokova, Luisa Meroni, and Mari Umeda, -. Somerville, MA: Cascadilla Proceedings Project. .
- Penka, Doris, and Hedde Zeijlstra. 2005. "Negative Indefinites in Dutch and German". Paper presented at the "20th Comparative Germanic Syntax Workshop", Tilburg.
- Percus, Orin. 2001. "Pragmatic Constraints on (Adverbial) (Temporal) Quantification". In *Papers in Linguistics 22: Papers on Predicative Constructions*, ed. G. Jaeger, A. Strigin, C. Wilder, and N. Zhang. Berlin: ZAS.
- Percus, Orin. 2006. "Antipresuppositions". In *Theoretical and Empirical Studies of Reference and Anaphora: toward the Establishment of Generative Grammar as an Empirical Science*, ed. A. Ueyama. Report of the Grant-in-Aid for Scientific Research (B), Project No. 15320052, Japan Society for the Promotion of Science, 52-73.
- Percus, Orin J. 1997. Aspects of 'A'. Doctoral Dissertation, MIT.
- Potts, Christopher. 2002. "No Vacuous Quantification Constraints in Syntax". In *Proceedings of NELS 32*, ed. Masako Hirotsu, 451–470. Amherst, MA: GLSA. .
- Potts, Christopher, Joe Pater, Rajesh Bhatt, and Michael Becker. 2008. "Harmonic Grammar with Linear Programming: From linear systems to linguistic typology". Ms, University of Massachusetts, Amherst; to appear in *Phonology*.
- Prince, Alan. 2002. "Entailed Ranking Arguments". ROA 500.
- Prince, Alan, and Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell. As Technical Report CU-CS-696-93, Department of Computer Science, University of Colorado at Boulder, and Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ, April 1993. Rutgers Optimality Archive 537 version, 2002.

- Prince, Alan, and Bruce Tesar. 2004. "Learning Phonotactic Distributions". In *Constraints in Phonological Acquisition*, ed. R. Kager, J. Pater, and W. Zonneveld, 245–291. Cambridge University Press.
- Revithiadou, A., and Marina Tzakosta. 2004. "Markedness hierarchies vs. positional faithfulness and the role of multiple grammars in the acquisition of Greek". In *Proceedings of GALA 2003 (Generative Approaches to Language Acquisition)*, ed. S. Baauw and J. Van Kampen, 377–388. Utrecht: LOT Occasional Series. .
- Riemsdijk, H. van. 1989. "Movement and Regeneration". In *Linguistic Theory and Dialectology*, ed. Benincá, P., -. Dordrecht: Foris.
- Riggle, Jason. 2007. "The Complexity of Ranking Hypotheses in Optimality Theory". *Computational Linguistics* 1:–.
- Roberts, Craige. 2003. "Uniqueness in Definite Noun Phrases". *Linguistics and Philosophy* 26:287–350.
- Russell, Benjamin. 2004. "Against Grammatical Computation of Scalar Implicatures". Paper presented at the workshop on Implicature and Conversational Meaning, August 2004, Nancy.
- Russell, Benjamin. 2006. "Against Grammatical Computation of Scalar Implicatures". *Journal of Semantics* 23:361–382.
- Sauerland, Uli. 2000. "No 'No': On the Crosslinguistic Absence of a Determiner 'No'". Paper presented at the "Tsukuba Workshop on Determiners and Quantification", Tsukuba University, Japan.
- Sauerland, Uli. 2002. "The Present Tense is vacuous". *Snippets* 6.
- Sauerland, Uli. 2003a. "A new semantics for number". In *Proceedings of SALT13*.
- Sauerland, Uli. 2003b. "An Integrated Theory of Agreement". Talk delivered at MIT on December 4.
- Sauerland, Uli. 2003c. "Implicated Presuppositions". In *Proceedings of the conference "Polarity, Scalar Phenomena, Implicatures"*. University of Milano Bicocca, June 18-20, 2003.
- Sauerland, Uli. 2004a. "A Comprehensive Semantics for Agreement". Paper presented at the "Phi-Workshop", McGill University, Montreal, Canada.
- Sauerland, Uli. 2004b. "On Embedded Implicatures". To appear on the *Journal of Cognitive Science* 5.
- Sauerland, Uli. 2004c. "Scalar Implicatures in Complex Sentences". *Linguistics and Philosophy* 27:367–391.
- Sauerland, Uli. 2007. "German Plural Negatives with Plural-Individual-Level Predicates". *Snippets* 15 March:6–7.
- Sauerland, Uli. 2008. "Implicated Presuppositions". In *Sentence and Context. Language, Context and Cognition*, ed. A Steube, -. Berlin, Germany: Mouton de Gruyter.
- Schlenker, Philippe. 2005. "The Lazy Frenchman's approach to the Subjunctive". In *Proceedings of Romance Languages and Linguistic Theory 2003*, ed. T. Geerts, I. van Ginneken, and H. Jacobs, 269–309. Benjamins.
- Schlenker, Philippe. 2006. "Maximize Presupposition and Gricean Reasoning". Manuscript, UCLA and Institute Jean-Nicod.

- Schmitt, Cristina. 2005. "Semi-Copulas. Event and Aspectual Composition". In *Aspectual Inquiries*, ed. P. Kempchinsky and R. Slabakova, 121–145. The Netherlands: Springer.
- Schubert, Lenhart K., and Francis Jeffrey Pelletier. 1989. "Generically Speaking, or Using Discourse Representation Theory to Interpret Generics". In *Properties, Types and Meaning*, ed. Gennaro Chierchia, Barbara Partee, and Raymond Turner, 193–268. Kluwer Academic Publisher.
- Schwarzschild, Roger. 1996. *Pluralities*. Dordrecht, Boston, London: Kluwer Academic Publisher.
- Shanon, Benny. 1976. "On the Two Kinds of Presuppositions in Natural Language". *Foundations of Language* 14:247–249.
- Sharvit, Yael, and Jon Gajewski. 2007. "On the Calculation of Local Implicatures". In *Proceedings of the 26th West Coast Conference on Formal Linguistics*, ed. C. Chang and H. Haynie, 411–419.
- Sharvy, R. 1980. "A More General Theory of Definite Descriptions". *The Philosophical Review* 89:607–624.
- Simons, M. 2000. *Issues in the Semantics and Pragmatics of Disjunction*. Garland Pub., New York and London.
- Simons, Mandy. 2006. "Notes on Embedded Implicatures".
- Singh, Raj. 2009a. "Maximize Presupposition! and Informationally Encapsulated Implicatures". In *Proceedings of SuB13*, ed. A. Riester and T. Solstad. Stuttgart.
- Singh, Raj. 2009b. MIT lecture notes.
- Smith, Jennifer. 2000. "Positional faithfulness and learnability in Optimality Theory". In *Proceedings of ESCOL99*, ed. Rebecca Daly and Anastasia Riehl, 203–214. Ithaca, New York: CLC Publications.
- Smolensky, Paul. 1996a. "On the Comprehension/Production Dilemma in Child Language". *Linguistic Inquiry* 27.4:720–731.
- Smolensky, Paul. 1996b. "The Initial State and Richness of the Base in Optimality Theory". John Hopkins Technical Report.
- Spector, Benjamin. 2006. *Aspects de la Pragmatique des Operateurs Logiques*. Doctoral Dissertation, Universite Paris 7.
- Spector, Benjamin. 2007a. "Aspects of the Pragmatics of Plural Morphology: On Higher-Order Implicatures". In *Presuppositions and Implicatures in Compositional Semantics*, ed. U. Sauerland and P. Stateval, 243–281. Palgrave-Macmillan.
- Spector, Benjamin. 2007b. "Higher Order Implicatures: plural vs. Number-Neutral Readings of Plural Indefinites". Talk presented at the Harvard Theory Group on the 5th of March, 2007.
- Spector, Benjamin. 2007c. "Scalar Implicatures: Exhaustivity and Gricean Reasoning". In *Questions in Dynamic Semantics*, ed. Maria Aloni, Alastair Butler, and Paul Dekker, 229–254. Elsevier.
- Stalnaker, Robert C. 2002. "Common Ground". *Linguistics and Philosophy* 25:701–721.
- Stemberger, Joseph Paul, and Barbara Handford Bernhardt. 1999. "The Emergence of Faithfulness". In *The Emergence of Language*, ed. B. MacWhinney, 417–446. Mahweh, NJ: Erlbaum.
- Stemberger, Joseph Paul, and Barbara Handford Bernhardt. 2001. "U-shaped learning in language acquisition, and restrictions on error correction". Poster presented at the 2001 Biennial Meeting of the Society for Research in Child Development.

- Stemberger, Joseph Paul, Barbara Handford Bernhardt, and Carolyn E. Johnson. 1999. "U-shaped learning in the acquisition of prosodic structure". Poster presented at the sixth International Child Language Congress.
- de Swart, H. 1991. *Adverbs of Quantification: A Generalized Quantifier Approach*. Doctoral Dissertation, University of Groningen. Published by Garland Press, New York, 1993.
- Tesar, Bruce. 1995. "Computational Optimality Theory". Doctoral Dissertation, University of Colorado, Boulder. ROA 90.
- Tesar, Bruce. 1998. "Error-Driven Learning in Optimality Theory via the Efficient Computation of Optimal Forms". In *Is the Best Good Enough? Optimality and Competition in Syntax*, ed. Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis, and David Pesetsky, 421–435. Cambridge, MA: MIT Press.
- Tesar, Bruce. 2008. "Output-Driven Maps". Ms., Rutgers University; ROA-956.
- Tesar, Bruce, and Paul Smolensky. 1998. "Learnability in Optimality Theory". *Linguistic Inquiry* 29:229–268.
- Tesar, Bruce, and Paul Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, MA: The MIT Press.
- van Valin, R. D. 1986. "An Empty Category as the Subject of a tensed S in English". *Linguistic Inquiry* 17:581–586.
- Vapnik, Vladimir N. 1998. *Statistical Learning Theory*. John Wiley and sons.
- Weijer, J. C. van de. 1997. Personal communication; reported in Levelt and van de Vijver (1998) and Boersma and Levelt (2000).
- Westerthal, Dag. 1984. "Determiners and Context sets". In *Generalized Quantifiers in Natural Language*, ed. J. van Benthem and A. ter Meulen, 45–71. Dordrecht: Foris.
- Wilkinson, Karina. 1991. *Studies in the Semantics of Generic Noun Phrases*. Doctoral Dissertation, University of Massachusetts, Amherst.
- Yatsushiro, Kazuko. 2005. "The Acquisition of Presuppositions: The Presuppositions of 'Every'". Talk presented at the workshop "Experimental Pragmatics: Exploring the Cognitive Basis of Conversation", University of Cambridge.