Advances in Optimality Theory
Editors: Ellen Woolford, University of Massachusetts Amherst, and Armin Mester, University of California, Santa Cruz

Optimality Theory is an exciting new approach to linguistic analysis that originated in phonology but was soon taken up in syntax, morphology, and other fields of linguistics. Optimality Theory presents a clear vision of the universal properties underlying the vast surface typological variety in the world's languages. Cross-linguistic differences once relegated to idiosyncratic language-specific rules can now be understood as the result of different priority rankings among universal but violable constraints on grammar.

Advances in Optimality Theory is designed to stimulate and promote research in this provocative new framework. It provides a central outlet for the best new work by both established and younger scholars in this rapidly moving field. The series includes studies with a broad typological focus, studies dedicated to the detailed analysis of individual languages, and studies on the nature of Optimality Theory itself. The series publishes theoretical work in the form of monographs and coherent edited collections as well as pedagogical texts and reference texts that promote the dissemination of Optimality Theory.

Forthcoming
Optimality Theory, Phonological Acquisition and Disorders
Edited by Daniel A. Dinnsen and Judith A. Gierut

Modeling Ungrammaticality in Optimality Theory
Edited by Curt Rice

Phonological Argumentation: Essays in Evidence and Motivation
Edited by Steve Parker

Understanding Allomorphy: Perspectives from Optimality Theory
Edited by Bernard Tranel

Hidden Generalizations

Phonological Opacity in Optimality Theory

John J. McCarthy

equinox
LONDON OAKVILLE

Published by
UK: Equinox Publishing Ltd., Unit 6, The Village, 101 Amies St., London SW11 2JW
USA: DBBC, 28 Main Street, Oakville, CT 06779
www.equinoxpub.com

First published 2007

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

ISBN-10	1845530519	(hardback)
	1845530527	(paperback)
ISBN-13	9781845530518	(hardback)
	9781845530525	(paperback)

Library of Congress Cataloging-in-Publication Data
McCarthy, John J., 1953-
Hidden generalizations : phonological opacity in optimality theory / John J. McCarthy.
p. cm. -- (Advances in optimality theory)
Includes bibliographical references and index.
ISBN 1-84553-051-9 (hb) -- ISBN 1-84553-052-7 (pb) 1. Grammar, Comparative and general--Phonology. 2. Opacity (Linguistics) 3. Optimality theory (Linguistics) I. Title. II. Series.
P217.3.M39 2006
414--dc22
2006009195

Typeset by Catchline, Milton Keynes (www.catchline.com)
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham

Contents

Acknowledgements

My colleagues and students at the University of Massachusetts were the first audience for this work and its predecessors, and they have given me much assistance along the way. I wish therefore to offer special thanks to Leah Bateman, Michael Becker, Tim Beechey, Della Chambless, Andries Coetzee, Paul de Lacy, Kathryn Flack, Maria Gouskova, Shigeto Kawahara, Mike Key, John Kingston, Elliott Moreton, Steve Parker, Joe Pater, Lisa Selkirk, Jen Smith, Anne-Michelle Tessier, Adam Werle, and Matt Wolf. I also had an opportunity to present this material in a short course at the University of Southern California. I am grateful to my hosts, Ania Łubowicz and Rachel Walker, to other participants from USC, and to Eric Bakovic, Bruce Hayes, Nicole Nelson, Jason Riggle, Colin Wilson, and Kie Zuraw. Their questions and comments were most valuable. In addition, I gave a lecture about this research at the Second Old World Conference in Phonology in Tromsø, Norway. For this opportunity and for much useful feedback, I thank the organizer, Curt Rice, and the other participants. Hamza Al-Mozainy could be counted on for speedy responses to my many emailed questions about Bedouin Arabic. Over the years, I have had many fruitful discussions of opacity and related topics with Alan Prince.

In addition to her other contributions, Kathryn Flack read the entire manuscript, pointing out weaknesses in the argumentation or exposition, correcting many inaccuracies, and generally making this work much better than it would otherwise have been.

I gratefully acknowledge the support of Armin Mester and Ellen Woolford, the editors of Advances in Optimality Theory. I am happy that this volume is the first of what will no doubt be a long line of works appearing in that series.

Readers are urged to consult my homepage, http://people.umass.edu/jjmccart, for any errata or additions to this text that came too late to make it into print.

This book is dedicated to the memory of my mother, Joan T. (Condon) McCarthy (June 9, 1928–November 17, 2005). 'For now we see as through a glass, darkly, but then face to face. Now I know in part, but then I will know even as I am known.'

1 Overview of the issues and the results

1.1 What is phonological opacity and why is it important?

The sound systems of languages are precisely that — systems. This means that they exhibit various regularities of structure. If phonology only dealt with those regularities that express categorical truths, however, it would not be a very interesting enterprise. In reality, phonological research and theory are to a great extent engaged with generalizations that fall short of categorical truth.

Some generalizations have lexical exceptions. Other generalizations are outmatched by requirements with higher priority. (See §2.3 for an example and discussion.) This book is about generalizations that are not quite true for a different reason: the truths they state are hidden by other aspects of the system. When a generalization is partially obscured in this way, it is said to be *opaque*. (The origin of this term is explained in §2.2.)

Since this book includes plenty of examples, a couple will suffice for now. In English, the past tense suffix is pronounced as [-əd] after [t] or [d]: *planted*, *braided*. This generalization is sometimes hidden, though, because [t] can be deleted after a nasal consonant: *planted* is pronounced carefully as [plæntəd] but more usually as [plænəd]. Even when the [t] is absent, the suffix is still pronounced as [-əd], so *planted* never merges with *planned*. Another example: some English dialects add a rising off-glide between [æ] and tautosyllabic [ŋ]: [bæe̯ŋ] *bang*. But when [ŋ] is derived by place assimilation from /n/, the generalization is hidden and does not seem to hold: [mæŋkɑjnd], *[mæe̯ŋkɑjnd] *mankind*.[1]

In these cases and others like them, generalizations of seemingly unquestionable validity turn out to be superficially invalid. I say 'superficially' because these generalizations are only invalid as naïve statements about surface structure: in the surface form [plænəd] *planted*, the past tense suffix is pronounced as [-əd] even though no [t] or [d] precedes, and in the surface form [mæŋkɑjnd] the [æ] has no off-glide even though [ŋ] follows. In a deeper sense, though, the generalizations really are valid, or at least they are invalid for principled reasons. The [ə] in [plænəd] is somehow a response to the underlying /t/ of *plant*, and the missing off-glide in [mæŋkɑjnd] is somehow a reflection of the /n/ that underlies [ŋ]. Furthermore, it is clear that these opaque generalizations are interacting with and being influenced by other regularities of the language, such as the loss of /t/ after /n/ in *planted* and the assimilation of /n/ to a following [k] in *mankind*.

A great deal of phonological research, ancient and modern, is devoted to understanding opacity. The earliest and most successful theory of opacity is the *derivation*. If the rule inserting [ə] before the English past tense suffix /-d/ precedes the rule deleting [t], then the [ə] rule sees a representation where the underlying /t/ is still present: /plænt-d/ $\rightarrow_{\text{ə-insertion}}$ [plæntəd] $\rightarrow_{t\text{-deletion}}$ [plænəd]. Likewise, if the rule inserting an off-glide between [æ] and [ŋ] is ordered before nasal place assimilation, then the glide-insertion rule sees a representation where the underlying /n/ is still present: [mænkɑjnd] $\rightarrow_{g\text{-epen}}$ *no change* $\rightarrow_{n\text{-assim}}$ [mæŋkɑjnd]. More recent work has explored alternatives to derivations, such as enrichments to phonological structure. This body of research is reviewed in §2.2 and §2.3.

It is not an exaggeration to say that the analysis of opacity has been one of the central themes of generative phonology. Although he does not frame it in these terms, opacity is the crux of Chomsky's (1964: 75ff.) argument against structuralist phonemics. He cites Joos's (1942) famous example from Canadian English, where the distinction between [ɹʌjɾəɹ] *writer* and [ɹɑjɾəɹ] *rider* is the result of derivational ordering: the rule that raises the /ɑj/ nucleus to [ʌj] before voiceless [t] is ordered before the rule that merges /t/ and /d/ into the voiced flap [ɾ] before an unstressed syllable. Opacity is also the main topic of the large and varied literature on rule ordering in generative phonology dating from about 1968 through 1980 (see §2.2). And opacity was at the heart of the controversy during that same period about the abstractness of underlying representations (see Kenstowicz and Kisseberth 1977: Chapter 1, 1979: Chapter 6). Abstract underlying forms only make sense when the rules that operate on them are opaque, since opaque rules are the only way for the abstract part of an underlying form to affect surface structure. In recent years, opacity has reemerged as an important challenge for phonological theories that rely primarily or exclusively on surface structure constraints, since opaque processes refer to conditions that are not visible in surface structure.

One reason for opacity's durability as an object of phonological research is that the alternatives to taking opacity seriously have proven to be unsatisfactory. Classical structuralist phonology was based on the premise that all authentic phonological generalizations are categorically true statements about the distribution of allophones. Chomsky's argument against structuralism therefore cites examples of generalizations that are not true in this sense,

echoing concerns that the structuralists themselves had already expressed (see, for example, many of the contributions to Joos 1957). Later work, in reaction to Chomsky and Halle (1968), took the position that phonological rules must state surface-true generalizations and must be unordered (see §2.2.3). If this were correct, then opaque processes would be nothing more than the lexicalized residue of sound changes that are no longer productive. It is difficult to accept that the productive, variable, low-level processes of English cited above are not part of speakers' active phonological competence, even though they are opaque. (Further evidence that processes can be productive but opaque is presented in §2.2.3 and §4.3.3.) The idea that phonological knowledge is reducible to surface-true generalizations has turned out to be an intellectual dead end. Phonological generalizations, whether they are formulated as rules or constraints, can be active and productive aspects of linguistic knowledge even if they are opaque.

All serious approaches to opacity are attempts to answer a single question: when not all phonological generalizations *can* be true, which ones *are* true and which are not? Derivations are one answer to this question: generalizations that hold of later stages in the derivation take precedence, truth-wise, over generalizations that obtain earlier in the derivation. For instance, the generalization that [ʌj] occurs before voiceless obstruents and [ɑj] before voiced ones is true early in the derivation, but not later on when /t/ and /d/ merge to the flap [ɾ] before an unstressed syllable. The early generalization about [ʌj] is not surface true, but the late generalization about [ɾ] is. Another example of an early generalization that ceases to be true because of a later one is the requirement that there be an off-glide between [æ] and tautosyllabic [ŋ]. Derivations, then, allow generalizations to state temporary truths. Phonological theories that require all generalizations to state durable truths, such as structuralism and Natural Generative Phonology, are unable to analyze opacity. Optimality Theory (Prince and Smolensky 2004) does not require generalizations to be surface true, and it deals with competition among generalizations through ranking. Ranking supplies a partial theory of opacity and perhaps, with the changes proposed in this book, a complete one.

1.2 What does this book have to say about opacity?

This book's principal thesis is that the best theory of opacity — and of phonology generally — is a synthesis of Optimality Theory (hereafter OT) with derivations. I argue that a candidate in OT includes not just a surface form but also a series of intermediate forms, each of which is minimally different from the form that immediately precedes it. A candidate, then, supplies information about the sequence of operations needed to link the underlying and surface forms. In the *planted* example, for instance, the winning candidate is the ordered *n*-tuple <plænt-d, plæntəd, plænəd>, and among its competitors is *<plænt-d, plænd>. In the *mankind* example, the winning candidate is <mænkɑjnd, mæŋkɑjnd>, and its most important competitor is *<mænkɑjnd, mæŋkɑjnd, mæe̯ŋkɑjnd>. These ordered *n*-tuples are called *candidate chains*. This theory is referred to as *OT with candidate chains*, or *OT-CC* for short. (On the antecedents of OT-CC, see §3.2.3.)

Candidate chains are subject to three well-formedness conditions, the details of which will be explained more fully in §3. First, all chains are *faithfully initiated*. This means that the first form in a chain is identical with the underlying representation, except for syllabification and the like.

Second, chains are *gradually divergent*. This condition has already been hinted at: the successive forms in a chain are minimally different from their neighbors, so the path from input to output proceeds in small steps.

Third, chains are *harmonically improving*. Every form in a chain is more harmonic than its predecessor, relative to the constraint hierarchy of the language in question. Because of the gradualness requirement a form's successor in a chain may not be the ultimate surface form, but it must be more harmonic.

The evaluation of a candidate chain by the grammar has some familiar properties: markedness constraints assess the last form in the chain, which is the chain's output; and faithfulness constraints measure discrepancies between the first and last forms in the chain. A novel type of constraint, Prec (for 'precedence'), specifies the preferred order of faithfulness violations in a chain. For example, the winning chain from the input /mænkɑjnd/ is opaque <mænkɑjnd, mæŋkɑjnd>. Its transparent competitor *<mænkɑjnd, mæŋkɑjnd, mæe̯ŋkɑjnd> loses because of the constraint Prec(Dep, Ident(Place)), which says that the Ident(Place)-violating mapping of /n/ to [ŋ] cannot precede the Dep-violating insertion of the off-glide [e̯].

This proposal raises two fundamental questions: Why analyze opacity by incorporating derivations into *OT*? Why analyze opacity by incorporating *derivations* into OT? These questions and the issues they raise merit serious consideration.

Why OT? Opacity figures prominently in critiques of OT by proponents of rule-based phonology (see §2.3). These critiques suffer from selective vision, however, seeing only OT's opacity problem and ignoring the very real problems that typically afflict rule-based phonology, such as dearth of explanation and absence of typological predictions. A central failing of rule-based phonology in the tradition of Chomsky and Halle (1968) is that it promotes descriptive completeness over all other goals of linguistic theory, a miscasting of priorities that means that the theory with the richest descriptive resources inevitably wins.

OT's very real advantages in explaining phonological systems and limiting their typological possibilities need not be reviewed here. (Readers who need convincing might want to consult McCarthy 2002b.) Since OT has many strengths, if opacity is OT's principal weakness, then it makes sense to explore ways of remedying this weakness instead of giving up on the enterprise.

Why derivations? There have been many proposals for accommodating opacity in OT, and I argue in §2.3 that most of them share a common characteristic: they rely on a third level of representation, neither underlying nor surface, as a crucial part of the analysis of opaque alternations. The defining property of a derivation, in the sense I am employing here, is the presence of this third (or fourth or fifth) level of representation. It is therefore not inaccurate to say that derivations or something like them have already been shown to play a necessary role in any reasonably complete approach to opacity in OT. But these more limited proposals for incorporating derivations into OT turn out to be insufficient when the full range of opaque alternations is considered, and so I argue in §3 that a richer theory of derivations, candidate chains, is required in OT.

If candidate chains are to be a welcome addition to OT, then they should offer more than just a way out of the opacity jam — and they do. For one thing, the well-formedness requirements on candidate chains have the effect of severely limiting the size of the candidate set. Because there is no natural limit on the number of epenthesis operations that Gen can perform, classic OT[2] imposes no upper bound on the length of a candidate and therefore no bound on the size of the candidate set for any input. But the harmonic improvement requirement ensures that candidate chains are of bounded length, for all inputs and for all constraint hierarchies (see §3.2.2).

For another, the harmonic improvement and gradualness requirements operate in concert to impose limitations on what kinds of mappings from underlying to surface representation are possible. In classic OT, a necessary and sufficient condition for the mapping /A/ → [B] in a language L is that [B] is more harmonic, according to L's constraint hierarchy, than any other candidate derived from /A/. With chains, this condition is still necessary but it is no longer sufficient: for [B] to be a possible output from /A/, there must be a well-formed chain connecting /A/ with [B]. Suppose that the chain must contain the intermediate form [C] because of the gradualness requirement: <A, C, B>. Then [C] must be more harmonic than [A] and less harmonic than [B] according to L's hierarchy. Sometimes this means that L's hierarchy must rank constraints that would be nonconflicting and therefore unrankable in classic OT. More importantly, sometimes it means that the /A/ → [B] mapping will be impossible in L. In this way, candidate chain theory restricts OT's power to perform certain global optimizations. This is also a major point of difference between chain theory and standard derivational models, which have no analogous notion of improvement or progress in their derivations (see §3.2.4.3).

The final chapter of this book presents case studies of two languages with significant amounts of opaque phonology. Because of my background and knowledge, both of the languages are varieties of Arabic, but I believe that they are fully representative of the kinds of opaque interactions that can be found in other, unrelated languages. Furthermore, most of the processes that are discussed are independent innovations rather than the legacies of a common ancestor. These case studies are a necessary adjunct to the theoretical proposal because many of the more complex issues in studying opacity only arise in analyses of sufficient depth to show more than two processes interacting (cf. Cathey and Demers 1970).

1.3 How should this book be read?

Readers who are new to OT are advised not to start with this book and to begin instead with a textbook introduction like Kager (1999a) or an overview like McCarthy (2002b). Either of those works will provide more than enough background to understand and critically evaluate the contents of the following chapters. The majority of readers will probably want to proceed linearly through §2 and §3, and then sample the extended analyses in §4. Those who are familiar with previous work on opacity before and since OT could skim rather than read §2. Readers who prefer praxis to theory may want to try reading §4 on the basis of just the brief introduction to candidate chains in §1.2, but I would not recommend it.

Notes

1 The *mankind* example comes from Donegan and Stampe (1979: 148–149). They also describe many other opaque interactions in English casual speech and dialect variation.

2 Throughout, I use the expression 'classic OT' as shorthand for the approach that has become a *de facto* standard, a synthesis of Prince and Smolensky's (2004) original proposals with correspondence-based faithfulness (McCarthy and Prince 1995, 1999). Among the characteristics of classic OT is a universal, finite constraint component CON that is limited to markedness and faithfulness constraints, and a single EVAL that evaluates fully formed output candidates that show the effects of all phonological processes in parallel.

2 Opacity, derivations, and Optimality Theory

2.1 Overview

This chapter begins (§2.2) by explaining what opacity is and how it is analyzed in rule-based phonology. The discussion then turns (§2.3) to a description of 'classic' Optimality Theory, the problems that opacity presents for classic OT, and various ideas about how to modify the classic theory to accommodate it. The conclusion I draw (§2.4) is that there is something fundamentally correct about rule-based phonology's serial derivation, leading to the proposal in §3 for an analogue of the serial derivation in a framework that retains all of classic OT's essential elements.

2.2 Opacity and derivations

2.2.1 Levels of representation

The theory of generative phonology recognizes two principal levels of representation, underlying and surface. At the underlying level, every morpheme has a unique representation. For example, the three principal surface alternants of the English plural suffix — [-z], [-s], and [-əz] — are derived by phonological rules from a single underlying representation, such as /-z/. Only suppletive or allomorphic alternants of morphemes require distinct underlying representations, such as the plural allomorphs /-ən/ of *children* and /-iː-/ of *geese*.

When a morpheme alternates nonsuppletively, its underlying representation must be discovered by the analyst and by the learner. In paradigms like German [bʊnt]/[bʊntə] 'multicolored/pl.' and [bʊnt]/[bʊndə] 'federation/pl.', distinct underlying representations are required because there are distinct patterns of voicing alternation: /bʊnt/ 'multi-colored' is voiceless throughout its paradigm and /bʊnd/ 'federation' alternates between voiced and voiceless. In theory and in actual practice, the relationship between the hypothesized underlying representation and the observed surface paradigm is sometimes less transparent than this.

Some recent research explores alternatives to positing an underlying level of representation. These approaches are monostratal in the sense that they recognize only a single level of representation, the surface form. In Declarative Phonology (Scobbie, Coleman, and Bird 1996), the work of underlying representations is done by constraints that describe morphemes. These descriptions are crucially incomplete in the case of alternating morphemes: e.g., for German [bʊnt]/[bʊndə], a constraint requires a final alveolar stop in 'federation' but says nothing about its voicing. Another monostratal approach seeks to express phonological generalizations purely in terms of relations between surface forms (e.g., Albright 2002, Burzio 2002).

In this context, it is worth reviewing the reasons why generative phonology posits an underlying level of representation (see Kenstowicz and Kisseberth 1979: chapter 6 for an overview of the evidence). The main argument comes from paradigms where the relationships among surface forms make sense only when mediated by an underlying form that is distinct from all of the surface forms. Schane's (1974) Palauan example in (2-1) is a well-known case. Because unstressed vowels reduce to [ə] and there is only one stress per word, disyllabic roots like 'cover an opening' and 'pull out' never show up with more than one surface nonschwa vowel. The hypothesized underlying representations /daŋob/ and /teʔib/ record the quality of the vowels as they appear when stressed in different members of the paradigm. These underlying representations incorporate all of the unpredictable phonological information about these morphemes. In generative phonology, the underlying representation of a root is the nexus of a set of related words, so it must contain sufficient information to allow the surface forms of all of those words to be derived by the grammar of the language. (See §4.3.3 for detailed argumentation in support of underlying representations in a case similar to Palauan.)

(2-1) Palauan Vowel Reduction

Underlying	Present Middle	Future Participle	
/daŋob/	mə-ˈdaŋəb	dəˈŋob-l	'cover an opening'
/teʔib/	mə-ˈteʔəb	təˈʔib-l	'pull out'

Generative phonology in the tradition of *The Sound Pattern of English* (SPE — Chomsky and Halle 1968) also allows for any number of levels intermediate between the underlying and surface levels. These intermediate levels are the result of sequential application of phonological rules. If a language has n rules in its grammar, it has n–1 intermediate representations, each of which is a potentially distinct way of representing the linguistic form that is being derived.

In Palauan, for example, the *SPE* theory requires an intermediate level at which stress has been assigned but vowel reduction has not yet applied: /daŋob-l/ $\rightarrow_{\text{stress}}$ [daˈŋobl] $\rightarrow_{\text{reduction}}$ [dəˈŋobl]. Indeed, *SPE* requires rules to apply sequentially even when simultaneous application would produce the same result.

2.2.2 Derivations

Any mapping from the underlying to the surface level of representation is a derivation. In this sense, any multistratal theory of phonology is derivational, including classic OT. The various multistratal theories differ significantly, however, in the complexity and internal organization of the derivations they posit.

The *SPE* approach to derivations retains considerable currency because it is often assumed even in contemporary research that has moved far beyond *SPE*'s other hypotheses about rules and representations (see §2.2.6). In *SPE*, the grammar consists of an ordered list of rules. The rules are applied in a strict sequence, with the output of rule i supplying the input to rule i+1. The output of each rule (except the last) is therefore a level of representation intermediate between the underlying and surface levels.

An important insight, due originally to Kiparsky (1968), is that rules may have different functional relationships to one another. In the least interesting case, a pair of rules may not interact at all — an example would be word-initial vowel epenthesis and word-final obstruent devoicing. When rules do interact, however, the functional relationship between them can often be classified as feeding or bleeding.

Rule A is said to *feed* rule B if A can create additional inputs to B. If A in fact precedes B, then A and B are in feeding order. (If B precedes A, then they are in counterfeeding order, which will be explained in §2.2.3.) An example of feeding order is the interaction between vowel and consonant epenthesis in Classical Arabic. Words that begin with consonant clusters receive prothetic [ʔi] (or [ʔu], if the next vowel is also [u]). As the derivation in (2-2) shows, prothesis of [ʔi] is the result of a feeding interaction between [i] epenthesis before word-initial clusters (= rule A) and [ʔ] epenthesis before word-initial vowels (= rule B).

(2-2) Feeding order in Classical Arabic

Underlying	/dˤrib/	'beat (m. sg.)!'
Vowel epenthesis	idˤrib	
[ʔ] epenthesis	ʔidˤrib	
Surface	[ʔidˤrib]	

Rule A is said to *bleed* rule B if A can eliminate potential inputs to B. If A in fact precedes B, then A and B are in bleeding order. (If B precedes A, then they are in counterbleeding order, which will also be explained in §2.2.3.) For example, in a southern Palestinian variety of Arabic, progressive assimilation of pharyngealization (= rule B) is blocked by high front segments, among them [i]. When the vowel [i] is epenthesized into triconsonantal clusters (= rule A), it also blocks assimilation, as shown in (2-3) (Davis 1995).

(2-3) Bleeding order in southern Palestinian Arabic

Underlying	/batˤn-ha/	'her stomach'
Vowel epenthesis	batˤinha	
Progressive assimilation	*Blocked*	
Regressive assimilation	bˤaˤtˤinha	
Surface	[bˤaˤtˤinha]	

Feeding and bleeding orders have something in common: when rules apply in feeding or bleeding order, those structures that are derived by rules are treated exactly the same as similar structures that were already present in underlying representation. For example, the process of [ʔ]-epenthesis in Classical Arabic applies to words with an underlying initial vowel, /al-walad-u/ → [ʔalwaladu] 'the boy (nominative)', and also to words with a derived initial vowel, such as the intermediate representation [idˤrib] in (2-2). Likewise, epenthetic and nonepenthetic [i] equally block progressive assimilation in Palestinian Arabic, as shown by (2-3) and /sˤiħħa/ → [sˤiħħa], *[sˤiˤħˤħˤaˤ] 'health'. In feeding and bleeding interactions, what you see is what you get: when derived and underived structures are identical, they exhibit identical phonological behavior. This is emphatically not the case with counterfeeding and counterbleeding interactions.

2.2.3 Opacity in derivations

If rule A feeds rule B and they are applied in the order B precedes A, then these rules are said to be in *counterfeeding* order. For example, in a Bedouin Arabic dialect (see §4.3.3), there are processes raising short /a/ to a high vowel in a nonfinal open syllable (= rule A) and deleting short high vowels in nonfinal open syllables (= rule B). These processes are in a feeding relationship, since raising has the potential to create new inputs to deletion. But their order is actually counterfeeding, as shown in (2-4). High vowels derived by raising are treated differently from underlying high vowels; only the underlying high vowels are subject to deletion. When rules apply in feeding order, derived and underlying structures behave alike, but when they apply in counterfeeding order, derived and underlying structures behave differently.

(2-4) Counterfeeding order in Bedouin Arabic[1]

Underlying	a. /dafaʕ/ 'he pushed'	b. /ʃarib-at/ 'she drank'
Deletion	—	ʃarbat
Raising	difaʕ	—
Surface	[difaʕ]	[ʃarbat]

The same is true of *counterbleeding* order, where rule A bleeds rule B but they are applied with B preceding A. In this same Arabic dialect, there is also a process palatalizing velars when they precede front vowels (see §3.3.3). Deletion (= rule A) bleeds palatalization (= rule B), since deletion can remove a high front vowel that would condition velar palatalization. But their order is counterbleeding, as shown in (2-5). High front vowels, even when they are absent from surface forms, induce adjoining velars to palatalize. Effects like this are typical with counterbleeding order.

(2-5) Counterbleeding order in Bedouin Arabic

Underlying	a. /ħa:kim-i:n/	b. /t-ħakum-in/
Palatalization	ħa:kʲimi:n	—
Deletion	ħa:kʲmi:n	tħakmin
Surface	[ħa:kʲmi:n]	[tħakmin]
	'ruling (masculine plural)'	'they (feminine) rule'

The result of counterfeeding and counterbleeding interactions is phonological opacity. Kiparsky's (1973: 79, 1976: 178–179) definition of opacity appears in (2-6). Clause (c) of this definition describes all processes of neutralization and so it is not relevant to our concerns here. We will therefore focus on clauses (a) and (b).

(2-6) Opacity

A phonological rule P of the form $A \rightarrow B / C\_\_D$ is *opaque* if there are surface structures with any of the following characteristics:

a. instances of *A* in the environment *C__D*,
b. instances of *B* derived by P that occur in environments other than *C__D*,
or c. instances of *B* not derived by P that occur in the environment *C__D*.

In the derivation /dafaʕ/ → [difaʕ] in (2-4), the high-vowel deletion rule is opaque under clause (a) of this definition: [difaʕ] has [i] (= *A*) in an open syllable (= *C__D*). Rules applied in counterfeeding order produce opacity of the clause (a) type, in which surface forms contain phonological structures that look like they should have undergone some process but in fact did not.[2]

In the derivation /ha:kim-i:n/ → [ħa:kʲmi:n] in (2-5), the palatalization rule is opaque under clause (b) of this definition: [ħa:kʲmi:n] has [kʲ] (= *B*) derived by palatalization (= *P*), but [kʲ] is not adjacent to a front vowel (= *C*__*D*). Rules applied in counterbleeding order produce opacity of this type, in which surface forms contain derived phonological structures without the context that shows how they were derived.

Counterfeeding and counterbleeding interactions supply the best — arguably, the only — evidence for language-particular rule ordering. It is not surprising, then, that skepticism about stipulated, language-particular ordering stimulated efforts to deny that opaque interactions involve living phonological processes (cf. §1.1). According to the proponents of Natural Generative Phonology (NGP), authentic phonological rules must state surface-true generalizations and they must be unordered (Hooper [Bybee] 1976, 1979, Vennemann 1972, 1974). NGP therefore maintains that opaque processes are merely the lexicalized residue of sound changes that are no longer productive — opaque rules were said to be 'not psychologically real'. (Recent work advocating similar views in an OT context includes Green (2004), Mielke, Hume, and Armstrong (2003), and Sanders (2002, 2003).) In fact, much if not all of the abstractness controversy of the 1970's, which dealt with proposed limits on the degree of disparity between underlying and surface representations (see Kenstowicz and Kisseberth 1977: Chapter 1, 1979: Chapter 6), was really an argument about opacity, since abstract underlying forms can influence the output only if opaque rules apply to them.

Certainly, there have been dubious analyses based on opaque rules and excessively abstract underlying forms, but outright denial of all opaque interactions is an empirically unsupportable overreaction. The example of Bedouin Arabic is instructive. (See §4.3.3 for detailed discussion.) Al-Mozainy (1981) presents several arguments that the opaque processes in this language are alive and productive. First, they are active in borrowed words. Second, high vowel deletion, even though it is opaque, applies productively in external sandhi, as shown in (2-7). If a process applies in external sandhi, it cannot be lexicalized, since it is impossible to list the infinite number of word collocations that the syntax provides.[3]

(2-7) Phrase-level deletion in Bedouin Arabic (Al-Mozainy 1981: 50–51)

/kaːtib al-ʒawaːb/	kaːt.bal.ʒu.waːb	'writing the letter'
	*kaː.ti.bal.ʒu.waːb	
/tiʕtˤuːnih al-museːʕiːdi/	tiʕ.tˤuːn.hal.m.seː.ʕiː.di	'you give it to the one
	*tiʕ.tˤuː.ni.hal.m.seː.ʕiː.di	from the clan of Musaiʿīd'

Third, the most compelling evidence that raising is productive comes from a kind of play language. Although raising usually affects any short /a/ in a nonfinal open syllable, there are phonological conditions under which raising regularly

fails to apply: after a guttural consonant ([ʔ], [h], [ʕ], [ħ], [χ], [ʁ]), or before a guttural consonant or coronal sonorant ([l], [r], [n]) that is itself followed by [a]. Bedouin Arabic has a secret language that permutes the consonants of the root, and this will sometimes affect the position of gutturals or coronal sonorants relative to the potentially raised vowel. When that happens, the vowel raises or fails to raise in exact conformity with these generalizations, as (2-8) shows. Other secret language data show that palatalization is also productive, even though it is opaque (see §3.3.3). In sum, the opaque phonology of Bedouin Arabic is also its living, productive phonology. (For further examples of processes that are productive yet opaque, see Donegan and Stampe (1979).)

(2-8) Raising alternations in a secret language

/dafaʕ/	Underlying representation
difaʕ	Unpermuted form
fidaʕ	Raising as expected
daʕaf	No raising before guttural + [a]
faʕad	"
ʕadaf	No raising after guttural
ʕafad	"

Although this sort of evidence shows that opacity is a fact of phonological life, certain types of opacity have received and deserve a skeptical reception. A famous example is *SPE*'s /ɹixt/ → [ɹɑjt] *right*. The point is that a few dubious analyses are not grounds to reject a theoretical construct, particularly when it is strongly supported by sound analyses, as it is in Bedouin Arabic.

A type of opacity that received particular attention in the 1970's is the Duke-of-York derivation (Hogg 1978, Pullum 1976). Like the eponymous Duke of the nursery rhyme,[4] underlying /A/ is changed by a rule to intermediate [B], but a later rule changes [B] back into [A]. Unlike the Duke's peregrinations, this activity is not as pointless as it seems: during the temporary [B] stage, erstwhile /A/ may opaquely escape an A-affecting process or cause a B-triggered one. More often, though, Duke-of-York derivations are simply an artifact of the commitment to sequential rule application. We will return to this topic, with exemplification, in §2.3.2.

2.2.4 Simultaneous application

Discussions of rule ordering often overlook an important alternative to the sequential derivation: simultaneous application. In many cases, rules could be applied simultaneously with no loss of generality, and so it is worth exploring which phenomena are and are not consistent with simultaneous application (for discussion, see Anderson 1974: 64–67, Donegan and Stampe 1979: 150, Hyman 1993: 204ff., Koutsoudas 1976, Koutsoudas, Sanders, and Noll 1974: 5–8).

Since simultaneous application is a somewhat unfamiliar notion, we should first get clear on what it means in rule-based phonology. A phonological rule describes a configuration that must be met in the rule's input — the rule's structural description — and a change that is to be effected in the rule's output — its structural change. If two rules are applied simultaneously, then their structural descriptions are analyzing exactly the same representation. It follows, then, that neither rule has access to any information that is contributed by the other rule's structural change. In sequential application, by contrast, the later rule always has access to information contributed by the earlier rule's structural change.

Opaque interactions are often compatible with simultaneous application, but transparent interactions require sequential application. The counterbleeding derivation /ħaːkim-iːn/ → [ħaːkʲmiːn] in (2-5), for example, would also work if the rules of palatalization and deletion were applied simultaneously. The structural description of the palatalization rule analyzes an input that contains [k] before [i], and so [k] is palatalized with complete indifference to the fact that the deletion rule is analyzing that same input toward the goal of deleting [i]. The important thing in this opaque derivation is that deletion must not precede palatalization; that desideratum could in principle be fulfilled by ordering palatalization before deletion, as in (2-5), or by requiring them to apply simultaneously.

Similarly, the counterfeeding derivation /dafaʕ/ → [difaʕ] in (2-4) is possible if deletion of high vowels and raising of low vowels apply simultaneously. The structural description of the high-vowel deletion rule is not met by /dafaʕ/, but the raising rule's structural description is met, so only raising actually applies. The important thing in this opaque derivation is that deletion should not apply to the output of raising; that desideratum could in principle be fulfilled by ordering raising before deletion, as in (2-4), or by requiring them to apply simultaneously.

Feeding and bleeding interactions, however, are incompatible with simultaneous application. In the feeding derivation /dˤrib/ → [ʔidˤrib] (2-2), for instance, the structural description of [ʔ] epenthesis is not met until after vowel epenthesis has applied, so sequential application is necessary. In the bleeding derivation /batˤn-ha/ → [bˤaˤtˤinha] (2-3), simultaneous application of vowel epenthesis and progressive assimilation would produce the result *[bˤaˤtˤinˤhˤaˤ], in which the epenthetic vowel is neither subject to nor a blocker of assimilation.

It is interesting that simultaneous application of rules typically produces opaque interactions but not transparent ones (unless the rules do not interact at all). Classic OT, though it evaluates candidates in which the effects of several processes are felt simultaneously, can model transparent interactions but not opaque ones (see §2.3.3). The reason for this difference is that rules and OT markedness constraints analyze different levels of representation. The structural description of a rule is met by the rule's *input*, which is sometimes identical to the underlying representation. The structural description of an OT markedness constraint is met in the ultimate *output*, the surface representation. Opacity requires reference to conditions obtaining in presurface representations, whereas transparency requires reference to conditions obtaining in surface representations.

2.2.5 Theories of rule ordering

In *SPE*, the order in which the rules are applied is *extrinsic*, which means that it is imposed on the rules by the language-particular grammar and cannot usually be predicted from rule form or function. From about 1969 through 1980, a voluminous literature developed around the question of whether some or even all aspects of rule ordering could be predicted. (See Anderson (1979: 15–18) and Iverson (1995) for brief surveys or Anderson (1974) and Kenstowicz and Kisseberth (1977: chapters 4, 6) for more extensive discussion.)

An *SPE*-style phonology of Classical Arabic must include a statement to the effect that vowel epenthesis precedes [ʔ] epenthesis to ensure that these rules apply in the order observed in (2-2). In some revisions of that model (e.g., Anderson 1974, Koutsoudas, Sanders, and Noll 1974), this ordering statement was dismissed as superfluous on the grounds that feeding order is unmarked or natural. In what sense is feeding order natural? If rules are allowed to apply freely at any point in the derivation when their structural descriptions are met, then the result will be the same as (2-2). Feeding orders maximize rule applicability. As was noted in §2.2.2, feeding orders also help to ensure that rules enforce true generalizations about surface structure: in Arabic, no word starts with a vowel because [ʔ] epenthesis is ordered after and thereby fed by vowel epenthesis. In the terminology of Donegan and Stampe (1979: 157), counterfeeding order of vowel epenthesis and [ʔ] epenthesis would act as a 'constraint' on the latter, preventing it from acting on derived representations and so rendering the [ʔ]-epenthesis generalization not categorically true.

Although feeding order was generally seen as natural and a second natural order was believed to exist, there is disagreement in the literature of that era over the question of whether the other natural rule order is bleeding or counterbleeding. In the earliest work on this topic, Kiparsky (1968) argued that historical change tends to maximize feeding and minimize bleeding orders. On the assumption the languages are attracted toward natural rule orders, this would mean that feeding and counterbleeding orders are natural. Anderson (1974) integrates this idea into his theory of local ordering, according to which feeding and counterbleeding order constitute a default case that can only be overridden by language-particular stipulation. Anderson's evidence includes analyses, none of them uncontroversial, in which maintaining the natural interaction between a pair of rules can cause them to apply in different orders within a single language. (This is the sense in which ordering is 'local': the theory comprehends ordering as a local relation between a pair of rules rather than a global list of ordered rules in the *SPE* fashion. Cf. §3.2.3.) Koutsoudas, Sanders, and Noll (1974) also argue for the naturalness of counterbleeding order.

Another body of work took the position that bleeding rather than counterbleeding order is natural (Iverson 1974, Kenstowicz and Kisseberth 1971, Kiparsky 1971). Apart from disagreements about analyses, the dispute is really one about the principle that determines the natural orders. If feeding and counterbleeding orders are natural, then what makes them natural is a principle that favors maximizing rule applicability: if A feeds B, then A supplies additional opportunities for rules to apply; and if A is not allowed to bleed B, then A cannot steal away some of B's opportunities to apply. If feeding and bleeding orders are natural, then what makes them natural is a principle that favors maximizing rule transparency: counterfeeding and counterbleeding orders produce opacity, whereas feeding and bleeding orders produce transparency, in which the effects of phonological generalizations are visible at surface structure.

In the course of research during the 1970's, these and other ordering principles were discussed, and there were even proposals about priority relationships among them (Anderson 1974: 217–218, Iverson 1976). The ultimate goal of the research program, according to some (e.g., Koutsoudas, Sanders, and Noll 1974), was the elimination of all language-particular ordering statements in favor of universal principles of applicational precedence. Supposedly *prima facie* arguments against this position have been adduced, such as two Canadian English dialects that differ solely in rule order (Bromberger and Halle 1989, Joos 1942), but in reality the argument is not that easy to make (Iverson 1995: 612–613). There never was a knock-down argument in support of language-particular ordering, nor was there general agreement on rule-ordering principles. Instead, the decade ended with a tacit consensus that research on universals of rule ordering had gone about as far as it could go.

2.2.6 Later developments

Interest in the topic of rule ordering waned around 1980. (An important exception is Goldsmith (1993b).) As the focus of phonological research moved elsewhere, however, matters of rule ordering and interaction sometimes reemerged in new contexts.

The development of nonlinear phonology and underspecification theory, beginning with works like Goldsmith (1976a), Kahn (1976), Liberman (1975), Liberman and Prince (1977), Clements and Ford (1979), McCarthy (1981), Prince (1983), and Archangeli (1984), took some of the analytic pressure off of phonological rules and shifted it to well-formedness constraints on phonological representations. In principle, an enriched theory of representations might lead to a reduction in the need for language-particular rule ordering, but in actual practice this line of research received little attention.

Satisfaction of representational constraints, however, required a new class of persistent rules that apply automatically at any point in the derivation when they are required (Chafe 1968, Myers 1991a). For example, when a consonant becomes unsyllabified in the course of a derivation, a persistent rule immediately adjoins it to a nearby syllable: /patika/ → $[pa]_\sigma\,[ti]_\sigma\,[ka]_\sigma \rightarrow_{\text{syncope}} [pa]_\sigma\, t\, [ka]_\sigma \rightarrow_{\text{persistent}} [pat]_\sigma\,[ka]_\sigma$. An important role of persistent rules, then, is to repair violations of well-formedness conditions, thereby ensuring that these conditions are respected not only at the beginning or end of the derivation but also in the middle. The free (re-)applicability of persistent rules is, of course, consistent with the principle favoring maximal rule application that was mentioned at the end of §2.2.5.

Another relevant post-1980 development is the theory of Lexical Phonology (Kaisse and Hargus 1993b, Kaisse and Shaw 1985, Kiparsky 1982, 1985, Mohanan 1982, among many others). Lexical Phonology is an extension of *SPE*'s theory of cyclic rule application. Certain rules may be designated as cyclic — in *SPE*, these are the English stress rules — and this causes them to apply repeatedly to successively larger morphological or syntactic constituents. The cycle accounts for transderivational similarities like the following:[5]

(i) Monomorphemic words like *ˌKalamaˈzoo* and *ˌWinnepeˈsaukee* exhibit the normal English stress pattern when three light syllables precede the main stress. Derived words like *acˌcrediˈtation* and *iˌmagiˈnation* deviate from this pattern under the influence of *acˈcredit* and *iˈmagine*.

(ii) A closed, sonorant-final syllable is normally unstressed in prestress position: *ˌserenˈdipity*, *ˌgorgonˈzola*, *ˌPennsylˈvania*. But the same kind of syllable may be stressed in the derived words *ˌauˌthenˈticity* and *ˌconˌdemˈnation* under the influence of *ˌauˈthentic* and *conˈdemn*.

In *SPE*, the aberrant stress of derived words is explained by their bracketing and cyclic application of stress. The stress rules first apply on the inner constituents of [*accredit*]*ation* or [*authentic*]*ity* and then on the outer constituents. The primary stress assigned on the first cycle becomes a secondary stress on the second cycle, when the stress rule reapplies and a new primary stress is assigned further to the right. Monomorphemic *Kalamazoo* and *serendipity* have no inner cycle, so they show the effects of just a single pass through the stress rules.

Lexical Phonology departs from *SPE* in regarding cyclic application as the norm rather than the exception for certain phonological rules. In addition, Lexical Phonology imposes further structure on the grammar, dividing the phonology up into separate components, called strata. At a minimum, there are two such strata, lexical and postlexical. The input to the lexical stratum is the underlying representation; the output of the lexical stratum is the input to the postlexical stratum; and the output of the postlexical stratum is the surface representation. Each stratum is a separate phonological grammar, though specific overlap requirements have sometimes been imposed (Borowsky 1986, Kiparsky 1984: 141–143, Myers 1991b) (see §2.3.4.2). It is usually assumed that the lexical stratum actually consists of several strata, and at each lexical stratum a different set of morphological and phonological processes may be in effect. For example, English suffixes like *-ity* are affixed in the first lexical stratum, and that is also where the stress assignment rules apply. Suffixes like *-ness* are not attached until the second lexical stratum, at which point the stress assignment rules are no longer active. That is why suffixes like *-ity* are stress-determining and suffixes like *-ness* are stress-neutral. It is sometimes also assumed that rules apply cyclically within each lexical stratum, as each affix of that stratum is added: e.g., /period/ → *ˈperiod* → *periˈodic* → *perioˈdicity*, all within the first lexical stratum.

Lexical Phonology retains *SPE*'s assumption that the rules within a grammar (= a stratum) are in a strict linear order. Despite this within-grammar strict ordering, the same rule can be observed to reapply at different points in the course of an entire derivation. As in *SPE*, the cycle offers one opportunity: a rule can reapply in the same stratum as multiple affixes are added. But even without any affixation at all, a rule can reapply if it is assigned to more than one stratum. (This situation is not unusual.) If a rule is included in the grammar of more than one stratum, it will simply reapply when the later stratum is reached. Lexical Phonology is thereby able to reanalyze some (perhaps all) of the evidence that had earlier been adduced in favor of a principle of unmarked feeding order or maximizing rule application. An example is Kiparsky's (1984) reanalysis of the interaction of Icelandic *u*-umlaut and syncope, which had previously been cited by Anderson (1974) as evidence for local ordering. Instead of allowing *u*-umlaut and syncope to apply in either order, whichever produces a feeding relationship, the Lexical Phonology approach fixes the within-stratum order as *u*-umlaut precedes syncope, but then allows *u*-umlaut to follow syncope by reapplying in a later stratum.

Assignment of rules to different strata offers a way of imposing extrinsic ordering on them: if rule A applies only in stratum 1 and rule B applies only in stratum 2, then A necessarily precedes B. Therefore, assignment of rules to strata could be used to reproduce some of the effects of *SPE*-style extrinsic ordering. This leads to some questions: Is extrinsic ordering within strata truly necessary? Could all rules in the same stratum apply simultaneously or in a universally predictable order? These questions were not asked, much less answered, in mainstream work on Lexical Phonology, though they were discussed in work that is not usually identified with the Lexical Phonology research program (Goldsmith 1993a, Lakoff 1993). In any case, the questions persist to this day, as we will see in §2.3.4.2.

Apart from these developments, the common consensus about rule ordering and opacity did not change very much in the period after 1980. Most phonologists, perhaps more from a lack of interest than strong conviction, continued to assume something like the *SPE* model of rule interaction.

2.3 Opacity in Optimality Theory

2.3.1 Properties of classic OT

This section is not intended as an introduction to or comprehensive overview of OT (for the former see Kager (1999a), and for the latter see Prince and Smolensky (2004) or McCarthy (2002b)). Rather, the goal is to review those aspects of OT that assume particular significance in the analysis of opacity.

In OT, a grammar of a language is a ranking of constraints. Ranking differs from language to language, so the ranking relation between any given pair of constraints is not generally predictable. Because language-particular ranking offers a way of accounting for language differences, it is reasonable (though not strictly necessary) to adopt the null hypothesis that the constraints themselves are universal and so are drawn from a universal constraint component, called Con. If Con is indeed universal, as is standardly assumed in classic OT, then it is fair to say that OT is an inherently typological theory of language. In other words, OT and a specific hypothesis about Con combine to predict all and only the possible grammars of human languages.

In classic OT, the constraints in Con are limited to two types: markedness constraints evaluate output forms, favoring some over others; and faithfulness constraints evaluate input-output mappings, favoring those mappings that maintain identity. Classic OT, in the sense employed here, also incorporates the assumption that the faithfulness constraints are formalized in terms of a correspondence relation between input and output forms (McCarthy and Prince 1995, 1999). Because the substantive properties of Con, particularly the markedness constraints, are largely unknown, empirical research in OT is mostly focused on developing a detailed picture of Con. Some of this research has led to proposals for constraint types that are neither markedness nor faithfulness, such as antifaithfulness (Alderete 2001a, 2001b) or morpheme realization (Kurisu 2001), but these ideas go well beyond the limits of what I am calling classic OT.

OT is inherently comparative. In the simplest case, the evaluative component Eval applies a language-particular constraint hierarchy to the task of comparing two possible outputs derived from a common input. Of these two outputs, called candidates, the more harmonic one is that which performs better on the highest-ranking constraint on which they differ.[6] The most harmonic or optimal candidate is the one that is more harmonic, in this sense, than any of its competitors.

Moreton (2003) has shown that classic OT entails a requirement of *harmonic improvement*. Assume that every candidate set contains at least one candidate that is fully faithful by virtue of obeying all of the faithfulness constraints in Con.[7] If the output of an OT grammar is not this fully faithful candidate, then it must be a candidate that is less marked than the fully faithful candidate relative to the language's constraint hierarchy. Moreton provides a formal proof of this result, but the intuition behind it is also clear: since a classic OT grammar has only markedness and faithfulness constraints, the only reason to violate a faithfulness constraint is satisfaction of a higher-ranking markedness constraint. Informally, you can stay the same or get better, but you can't get worse.

2.3.2 Process interaction in classic OT

Except for digressions in chapter 2 of Prince and Smolensky (2004) and in the appendix of McCarthy and Prince (1993b), classic OT has usually included an assumption of *parallelism*. This means that the candidates under evaluation can show the effects of several phonological processes simultaneously — that is, the effects of processes are evaluated in parallel. Classic OT is therefore a bistratal theory: it recognizes two levels of representation, input and output, but

nothing in between. This is obviously very different from *SPE*, which has nearly as many intermediate levels of representation as there are rules (see §2.2.1).

In general, transparent interaction of processes is fully compatible with parallelism. Consider first a feeding interaction like (2-2), where underlying /dˤrib/ becomes surface [ʔidˤrib], showing the effects of two processes, vowel epenthesis and [ʔ] epenthesis. Taken separately, each process involves, *inter alia*, a basic markedness-dominates-faithfulness ranking, as shown in (2-9) and (2-10).[8] Faithful syllabification of the initial cluster in /dˤrib/ is impossible because of *Complex-Onset and other markedness constraints. Violation of the lower-ranking antiepenthesis constraint Dep is the chosen alternative. Faithful syllabification of a word-initial vowel is a breach of Onset, which also ranks above Dep. The feeding interaction between the two types of epenthesis is simply the result of satisfying both *Complex-Onset and Onset simultaneously. Among the candidates derived from /dˤrib/ is one in which both vowel and [ʔ] epenthesis have occurred. This candidate is favored by both of the high-ranking constraints, as tableau (2-11) illustrates.

(2-9) *Complex-Onset >> Dep

/dˤrib/	*Comp-Ons	Dep
→ ʔidˤrib		2
dˤrib	W_1	L

(2-10) Onset >> Dep

/al-walad-u/	Onset	Dep
→ ʔalwaladu		1
alwaladu	W_1	L

(2-11) Feeding interaction

/dˤrib/	Onset	*Comp-Ons	Dep
→ ʔidˤrib			2
dˤrib		W_1	L
idˤrib	W_1		L_1

I noted in §2.2.4 that transparent interactions are incompatible with simultaneous application of phonological rules. That is because a rule's structural description analyzes that rule's input, and the fed rule's structural description cannot be met until after the feeding rule has applied. In OT, however, the structural descriptions of markedness constraints analyze outputs, and feeding interactions are simply a consequence of satisfying such constraints. This is the sense in which classic OT exhibits parallelism: high-ranking markedness constraints can favor a candidate that differs from the input by the simultaneous effects of two or more processes, as in (2-11).

The situation is the same with the other type of transparent interaction, bleeding. The difference is that bleeding interactions may involve conflict between markedness constraints. For instance, the mapping /batˤn-ha/ → [bˤaˤtˤinha] in (2-3) shows that the markedness constraint responsible for progressive assimilation of pharyngealization is crucially dominated by two *other markedness constraints*, one forbidding pharyngealization of [i] and the other ruling out medial triconsonantal clusters (and thereby demanding [i] epenthesis). In this way, the output [bˤaˤtˤinha] is favored over alternatives like *[bˤaˤtˤiˤnˤhˤaˤ], with pharyngealized [i], and *[bˤaˤtˤnˤhˤaˤ], with a triconsonantal cluster.

Parallel evaluation in classic OT also eliminates the need for certain kinds of Duke-of-York derivations (see §2.2.3). An example comes from Nuuchahnulth, formerly known as Nootka (Campbell 1973, Kenstowicz and Kisseberth 1977: 171ff., McCarthy 2003c, Sapir and Swadesh 1978).[9] This language has a process that rounds velars and uvulars when they follow round vowels (2-12), as well as a process that unrounds velars and uvulars at the end of a syllable (2-13). (Syllable boundaries are shown by a period/full stop.) These two processes are in a mutual feeding relationship: when a velar or uvular follows a round vowel, as in (2-14), rounding creates inputs to unrounding and unrounding creates inputs to rounding. In the *SPE* tradition, this kind of conflict can only be resolved by rule ordering, and indeed mutual feeding relationships presented special challenges to those seeking to predict rule ordering (§2.2.5). The stipulated ordering is given in (2-14). Because unrounding gets its hands on the form later in the derivation, it states the surface-true generalization that syllable-final consonants are unrounded. The truth of the rounding generalization consequently suffers: there exist some nonrounded velars and uvulars that are preceded by a round vowel.

(2-12) Rounding in Nuuchahnulth

Underlying	/ħaju-qi/	'ten on top'
Rounding	ħa.ju.qʷi	
Surface	[ħa.ju.qʷi]	(cf. [hi.ta.qi] 'on top')

(2-13) Unrounding

Underlying	/ɬa:kʷ-ʃitɬ/	'to take pity on'
Unrounding	ɬa:k.ʃitɬ	
Surface	[ɬa:k.ʃitɬ]	(cf. [ɬa:.kʷiq.nak] 'pitiful')

(2-14) Duke-of-York derivation

Underlying	/mˀu:q/	'throwing off sparks'
Rounding	mˀu:qʷ	
Unrounding	mˀu:q	
Surface	[mˀu:q]	(cf. [mˀo.qʷak] 'phosphorescent')

In OT, deriving [mˀu:q] from /mˀu:q/ does not require passing through the intermediate step [mˀu:qʷ]. Rather, this is a matter of conflict between markedness constraints, and it is resolved, as are all constraint conflicts, by ranking the conflicting constraints. In (2-15), I introduce two ad hoc markedness constraints and show how the higher-ranking constraint is the one that favors nonround consonants syllable-finally. Both are ranked above the faithfulness constraint IDENT(round), to account for the predictability of consonant rounding in this context.

(2-15) $*K^{w}]_{\sigma}$ >> *uK >> IDENT(round)

/mˀu:q/	$*K^{w}]_{\sigma}$	*uK	ID(round)
→ mˀu:q		1	
mˀu:qʷ	W_1	L	W_1

It is useful to compare the *SPE*-style analysis in (2-14) with the OT analysis in (2-15). The comparison shows why parallelism is and should be the null hypothesis for OT. In the *SPE* model, ordering is a way of establishing priority relationships among rules, and in a case like Nuuchahnulth it is the last rule that has priority in the sense that it states a surface-true generalization, even though the earlier rule does not. In OT, priority relationships among constraints are established by ranking them, and this example shows that ranking can replace at least some applications of rule ordering. The null hypothesis, then, is that OT can dispense with ordering and all of its trappings, including intermediate derivational steps. In its place, OT has constraint ranking, which is required independently. This very strong claim is certainly not uncontroversial, and opacity presents the main challenge.

Before we go on to look at opacity in OT, however, it is appropriate to examine some conceptual arguments that have been advanced against parallelism and in favor of *SPE*-style serial derivations. One of these conceptual arguments holds that sequential rules accurately model a system of mental computation (Bromberger and Halle 1997). The failure of the Derivational Theory of Complexity showed that this idea is very far off the mark, at least in syntax (Fodor, Bever, and Garrett 1974); the same is true in phonology (Goldsmith 1993b). Indeed, if the goal of generative grammar is to construct competence models (Chomsky 1965), then it is a category mistake to ask whether these models faithfully replicate mental computation.

Another argument offered in favor of sequential rule application is that it makes sense in terms of language history (Bromberger and Halle 1989): the ordering of synchronic rules matches the chronology of diachronic sound changes. The principal problem with this view is that it misconceives language change. If language learners in generation Y innovate a sound change, they do not simply add a rule onto the end of generation X's phonological grammar — they cannot, since generation Y does not have direct access to generation X's internalized grammar. Generation Y's learning is informed exclusively by X's actual productions, as filtered through Y's perceptual system. X's productions offer only indirect evidence of X's grammar, subject to well-known limitations like the absence of negative evidence. From this perspective, we neither expect nor do we necessarily observe that grammars change by accreting rules at the end of the ordering.

2.3.3 Opacity in classic OT[10]

Classic OT recognizes just two types of constraints, markedness and faithfulness, and just two levels of representation, underlying and surface. Markedness constraints can refer to only one of those levels of representation, surface structure. Faithfulness constraints refer to both levels, but they can only do one thing: require identity. The standard derivational approach to opacity relies on having intermediate levels of representation (see §2.2.3), but classic OT has none. Furthermore, the limitation of markedness constraints to evaluating surface structure has unwelcome consequences for the analysis of counterbleeding opacity.

In counterbleeding opacity, a phonological process occurs even though the conditioning environment is not present in surface structure. In the Bedouin Arabic example (2-5), for instance, /k/ palatalizes even though it is not followed by a front vowel in the surface form: /ħa:kim-i:n/ → [ħa:kʲmi:n]. In other words, the /k/ → [kʲ] unfaithful mapping is a response to phonological

conditions that are not visible in the output form, though they are visible in the input. Because markedness constraints are limited to evaluating outputs, the markedness preference for [kʲ] over [k] before front vowels cannot be invoked to explain why /k/ is palatalized before a vowel that is no longer present. The problem is apparent from tableau (2-16), which shows that [ħa:kʲmi:n] is harmonically bounded by *[ħa:kmi:n].

(2-16) Counterbleeding opacity in classic OT[11]

/ħa:kim-i:n/	**i*CV	MAX	**ki*	ID(back)
→ ħa:kʲmi:n		1		1
a. ħa:kmi:n		1		L
b. ħa:kʲimi:n	W_1	L		1
c. ħa:kimi:n	W_1	L	W_1	L

Row (a) in (2-16) contains no W's and one L, so the candidate in (a) harmonically bounds the intended winner. Moreover, since this candidate is more faithful and less marked than the intended winner, no other classic OT faithfulness or markedness constraint could be introduced to break this harmonic bounding. (On an alternative analysis with coalescence, see §2.3.4.1. For the OT-CC analysis of palatalization, see §3.3.3, and for the analysis of syncope — minus the ad hoc constraint **i*CV — see §4.3.3.)

In this and other cases of counterbleeding opacity, an unfaithful mapping occurs for reasons that cannot be explained with classic OT markedness constraints because the conditions that encourage the unfaithful mapping are no longer apparent in surface structure. Although analyses of particular instances of counterbleeding opacity (including this one) have been proposed, there is no general solution that remains within the strictures of classic OT.

In contrast to counterbleeding opacity, counterfeeding opacity can in principle be accommodated in classic OT. Consider the Bedouin Arabic example in (2-4), in which underlying /i/ deletes (/ʃarib-at/ → [ʃarbat]) but [i] derived from /a/ does not (/dafaʕ/ → [difaʕ], *[dfaʕ]). As above, let **i*CV stand for the constraint that favors [ʃarbat] over faithful [ʃaribat]; it dominates MAX. Let **a*CV stand for the constraint that favors [difaʕ] over faithful [dafaʕ]; it dominates IDENT(+low). Tableau (2-17) shows that the desired output [difaʕ] is unattainable with just these four constraints. To circumvent this paradox, we require a constraint that favors the desired winner in (2-17) over the loser in (a). This constraint, which can be called MAX-A, forbids the /a/ → Ø mapping. MAX-A meets the formal requirements for faithfulness constraints: it requires identity between underlying and surface structure. Ranked above **i*CV, MAX-A correctly favors [difaʕ], as shown in (2-18). Furthermore, MAX-A does not interfere with the analysis of high vowel syncope in forms like /ʃarib-at/. (See §4.3.3 for the full analysis.)

(2-17) Impossibility of [difaʕ] without MAX-A

/dafaʕ/	**a*CV	ID(low)	**i*CV	MAX
→ difaʕ		1	1	
a. dfaʕ		L	L	W_1
b. dafaʕ	W_1	L	L	

(2-18) Counterfeeding opacity in classic OT

/dafaʕ/	MAX-A	**a*CV	**i*CV	ID(low)	MAX
→ difaʕ			1	1	
a. dfaʕ	W_1		L	L	W_1
b. dafaʕ		W_1	L	L	

In theory, this mode of analysis could be generalized to all instances of counterfeeding opacity, thereby providing classic OT with a ready-made solution to this half of the opacity problem. In practice, though, that would not be a good idea. Dealing with the full range of counterfeeding interactions will require a very rich faithfulness theory, undoubtedly much richer than we want or would otherwise need. Another counterfeeding interaction in Bedouin Arabic illustrates. Raising of /a/ to [i] in an open syllable is not fed by a process of epenthesis that breaks up final consonant clusters: /gabr/ → [gabur], *[gibur] 'grave'. To analyze this phenomenon in the same manner as (2-18), we would need a faithfulness constraint with the following definition: 'Assign a violation mark for every instance of a surface high vowel that stands in correspondence with an underlying low vowel, provided that this surface high vowel is followed in the next syllable by a vowel that has no underlying correspondent.' In other words, the counterfeeding interaction in [gabur] requires a version of IDENT(+low) that is applicable only if the vowel in the next syllable is

epenthetic. Constraints like this are necessarily embedded in a faithfulness theory that makes unattested and implausible typological predictions. Rather than demonstrate this now, I return to the matter in §2.3.4.1 when I discuss a theory of faithfulness that countenances such constraints, local conjunction.

Classic OT has an inherent bias toward transparent interactions (§2.3.2). Counterfeeding opacity requires undesirable enrichment of faithfulness theory, and counterbleeding opacity is usually intractable. Since opacity appears to be an authentic property of phonological systems, classic OT needs to be modified. The question is how.

2.3.4 Previous approaches to opacity in classic OT

More than a few different proposals have been made about how to integrate the analysis opacity into OT. Some are recent or short-lived; others date back to the earliest work in the theory. For discussion purposes, they can be grouped into four broad categories:

i) Changes in substantive properties of phonological representation or the constraint component CON (§2.3.4.1). The goal is to analyze some or all cases of opacity by enriching representations or creating new constraints. (The approach discussed at the end of the previous section is an example.)

ii) Introduction of intermediate derivational stages and something like rule ordering to OT (§2.3.4.2).

iii) Introduction of an equivalent of intermediate derivational stages, but without any direct counterpart to rule ordering (§2.3.4.3).

iv) Reinterpretation of opacity as a mechanism for preserving underlying contrasts (§2.3.4.4).

2.3.4.1 Opacity via novel substantive assumptions

This section describes approaches to opacity that place the main analytic burden on assumptions about substantive matters. Three lines of attack will be discussed in turn: *representational approaches*, which enrich surface structure in ways that allow opaque processes to be reanalyzed as transparent; reanalysis of counterbleeding opacity as a type of *segmental coalescence*; and reanalysis of counterfeeding opacity as a faithfulness effect using *local constraint conjunction*.

Representational approaches to opacity. Opacity issues arose in the very first work on OT, Prince and Smolensky (2004). The topic comes up in the context of two analyses, Lardil (pp. 145, 148) and Fula (p. 255).

In Lardil nominative case forms, final vowels are deleted (Hale 1973): /jilijili/ → [jilijil] 'oyster species (nominative)' (cf. the nonfuture accusative [jilijili-n], with suffix /-n/ and no truncation). When this apocope process exposes a final consonant that is not allowed syllable-finally (Wilkinson 1988), the consonant deletes as well: /ŋawuŋawu/ → [ŋawuŋa] 'termite'. Apocope therefore *feeds* consonant deletion. Crucially, apocope must not be fed by consonant deletion; if it were, then we would expect to find apocope and consonant deletion chewing through words until a licit coda is found (as in *[murkun] from /murkunima/ 'nullah') or the bimoraic word minimum is reached (as in *[kuru] from /kurumpuwa/ 'tata-spear'). This is an example of counterfeeding opacity.

Fula has two processes that refer to geminate consonants. One process shortens a geminate after a long vowel, and the other hardens geminate continuants into stops (Paradis 1988). They interact in counterbleeding fashion, with an underlying geminate continuant undergoing hardening even if it is also shortened: /la:w:i/ → [la:bi] 'roads'.

The analytic strategy that Prince and Smolensky apply to both of these cases is closely connected with their implementation of faithfulness constraints. Faithfulness is essential to OT, since without faithfulness markedness runs amok, driving every input down to some least marked output like [ba] (cf. Chomsky 1995: 380fn.). The idea of faithfulness is thus a key insight without which OT would be a failed enterprise. The implementational details are much less central, though relevant to the analysis of opacity. The implementation adopted in Prince and Smolensky (2004) is based on a principle dubbed Containment in McCarthy and Prince (1993b): all of the phonological material in the underlying representation must be preserved in every candidate output form.

Containment therefore entails that there are no literal deletion processes. Instead, the effects of deletion are obtained from the joint action of three additional assumptions, all with precedents elsewhere:

(i) Underlying representations lack prosodic structure, particularly syllabification.

(ii) Phonological material may remain unincorporated into prosodic structure.

(iii) Unincorporated phonological material receives no phonetic interpretation.

Thus, a deleted segment like the final /i/ of [jilijil] is present in the output form but syllabically unparsed: [ji]$_\sigma$ [li]$_\sigma$ [jil]$_\sigma$ i, or more compactly [jilijil<i>]. With a shortening process like /la:w:i/ → [la:bi], an underlying mora is preserved in the output but also syllabically unparsed. Both situations violate constraints from the Parse family, which require segments, moras, and other structural elements to be incorporated into prosodic structure. Some theories of syntactic deletion are a close parallel (e.g., Chomsky 1995).

Containment supplies an analytic strategy for many cases of opacity. For example, in Prince and Smolensky's analysis of Lardil, apocope is the result of satisfying the constraint Free-V 'Word-final vowels must not be parsed (in the nominative)' (p. 123). In [ŋawuŋa<wu>], the word-final vowel is unparsed, as requested, and the preceding [w] is unparsed because it is not a licit syllable coda. Nonparsing of the preceding [a], however, would violate Parse for no reason — if 'word-final' means 'rightmost segment, parsed or not', then the word-final vowel is [u], and [a] has no claim to word-final status. Apocope cannot feed itself, then, because apocope can only affect a vowel that is word-final in underlying representation. (This also explains why the last but not word-final vowel of /wuŋkunuŋ/ does not apocopate: [wuŋkunu<ŋ>] 'queen-fish'.)

In Fula, we need to explain how a mapping with hardening and degemination of geminate continuants (/la:w:i/ → [la:bi]) can be more harmonic than a mapping with degemination alone (/la:w:i/ → *[la:wi]). Since *[la:wi] is not pronounced with a geminate, it should not lose to the hardened former geminate in [la:bi]. Prince and Smolensky's solution (p. 255) again relies on Containment. The two skeletal positions linked to geminate /w/ in /la:w:i/ can never be literally deleted; rather, both are present but one is syllabically unparsed in candidates with degemination like [la:bi] and *[la:wi]. The markedness constraint against geminate continuants defines a 'geminate' as a consonant linked to two skeletal positions, regardless of whether the skeletal positions are syllabified. This markedness constraint, then, is sensitive to the representation and not the pronunciation, and the representation of *[la:wi] contains a 'geminate' continuant because it is derived from /la:w:i/ under Containment. Once a geminate, always a geminate, as far as this constraint is concerned.

This theory of opacity requires no changes in OT proper, and that makes it attractive. It has empirical problems of two types, however: there are observed opaque interactions that it cannot easily accommodate; and there are transparent interactions that ought to be opaque if this theory is right. We will examine each in turn. (For related discussion, also see the critique of Containment in McCarthy and Prince (1995, 1999).)

A basic prediction of the Containment model is that syllabification always interacts transparently with processes because syllabification is present only in the output. With syllabification though not with segmental structure, the pronunciation and the representation are true to one another. Cases like the Bedouin Arabic /gabr/ → [gabur] example (§2.3.3) are therefore problematic: /a/ raises to [i] in an open syllable, and [ga.bur] has an open syllable. There is no earlier stage of syllabification to refer to opaquely, in which /a/ is in a closed syllable. Another example along the same general lines can be found in Levantine Arabic (see §4.2). When a final cluster is resolved by epenthesis, stress is assigned to the erstwhile final syllable, in conformity with the general pattern for words ending in such 'superheavy' syllables: /katab-t/ → [ka'tabit] 'I wrote'. Forms with the same surface syllable structure but without epenthesis are stressed differently: /katab-it/ → ['katabit] 'she wrote'. Since syllabification is necessarily transparent under Containment, this opaque interaction between stress and syllabification/epenthesis is inexpressible.

The other problem is that many transparent processes, which should be unremarkable, end up tripping over the unparsed remnants of deletion (cf. Beckman 1997: 27–31). In Maltese, for example, there is a completely transparent process of regressive voicing assimilation in obstruent clusters (Borg 1997). Because it is completely transparent, this process also affects consonant clusters that are created by syncope: /ni-ktib-u/ → ['nigdbu] 'we write'. Under Containment, syncope does not affect string-adjacency relations among segments because no segment is literally deleted. Therefore, voicing assimilation affects a sequence of noncontiguous consonants: ['nigd<i>bu]. This is a surprising result, since voicing assimilation has never been observed to traverse a pronounced vowel in any language. This problem could be avoided by adopting a more sophisticated theory of locality that reckons segments as adjacent if no parsed segment appears between them, but this move would be inconsistent with the Containment-based analysis of Lardil, where unparsed segments do count in determining whether a vowel is final or not.

A usual (though not essential) accompaniment to Containment is the assumption that epenthesis is not literal segmental insertion but rather prosodic overparsing (after Broselow 1982, Ito 1986, 1989, Lowenstamm and Kaye 1986, Piggott and Singh 1985, Selkirk 1981b and others). In overparsing, syllables are created with empty structural positions. The phonetic content of these empty positions is determined extrasystemically — that is, outside the phonological grammar proper. Those positions that are devoid of segmental content violate faithfulness constraints from the FILL family, which militate against such mismatches between segmental and prosodic structure. An example: the phonological output corresponding to Classical Arabic [ʔidˤrib] is [ONdˤrib], where O and N stand for an unfilled onset and nucleus, respectively. The spell-out of O as [ʔ] and N as [i] happens in some later module that interprets the output structures derived by the OT phonological grammar.

Because the phonetic identity of epenthetic segments is supplied extraphonologically, processes of segmental phonology should treat them opaquely, as if they were not present, whereas syllable-sensitive processes should treat epenthesis transparently, for reasons already given. There are indeed some cases where epenthesis interacts opaquely with segmental phonology. For example, Herzallah (1990: 109–110) reports for her northern Palestinian Arabic dialect that the vowel [i] causes a preceding pharyngealized /rˤ/ to lose its pharyngealization: [ʔɑfrˤɑz] ~ [jifriz] 'he classified ~ he classifies'. Epenthetic [i] does not have this effect, however: /fɑrˤm/ → [fɑrˤim] 'cutting'. This observation is consistent with the claim that information about the quality of epenthetic vowels is determined after the phonological grammar has done its work. On the other hand, the example in (2-3) shows for a southern Palestinian dialect that epenthetic [i] blocks the spread of pharyngealization, acting just like non-epenthetic [i] in this respect. So epenthesis does not show consistent opaque interaction with segmental processes. Rather, interaction may be transparent or opaque on a language-specific basis. This is contrary to the predictions of the FILL-based model of epenthesis.

The Containment theory of faithfulness is, as we have seen, also a theory of opacity, but not an entirely successful one. Two main problems have been identified. Under Containment, deleted segments should be consistently visible to processes that are conditioned purely by segmental adjacency but consistently invisible to processes that are conditioned by syllable structure. This predicts opaque interactions in the former case and transparent interactions in the latter, but there are counterexamples to both predictions. Under the empty-node theory of epenthesis, epenthetic segments should be consistently invisible to processes that are conditioned by segmental adjacency but consistently visible to processes that are conditioned purely by syllable structure. This predicts opaque interactions in the former case and transparent interactions in the latter, but again there are counterexamples to both predictions. The inherent simplicity and consequent attractiveness of this theory of opacity yields to its empirical inadequacies.

There is some later work exploring enhancements of this theory of opacity to grant it greater descriptive power (Goldrick 2000, Goldrick and Smolensky 1998). The key idea of this approach, called Turbidity, is that the symmetric *is associated with* relation between prosodic and segmental structure is divided into two asymmetric relations: segments *project* prosodic structure, and prosodic structure *is pronounced as* segments. Usually, these two relations operate in tandem, with segment *s* projecting prosody *p* if and only if *p* is pronounced as *s*.

The hallmark of opacity in Turbidity theory is a mismatch between the Project and Pronounce relations. Compensatory lengthening presents a typical example.[12] Compensatory lengthening is a type of counterbleeding opacity: a deleted segment projects a mora, but that mora is pronounced with a different segment, thereby lengthening it. In Turkish, for example, coda /h/ is optionally deleted before a continuant or nasal, in which case the preceding vowel lengthens (Sezer 1985: 230): [kahve] ~ [ka:ve] 'coffee'. The representation of [ka:ve] is shown in (2-19), with upward and downward arrows standing for the Project and Pronounce relations, respectively. In this representation, the segments [a] and [h] each project a mora (upward arrows), but the mora projected by [h] is pronounced as [a] (diagonal downward arrow).

(2-19) Compensatory lengthening in Turbidity theory

```
 μ μ   μ
 ↕↙↑   ↕
k a h v e
```

In Turbidity theory, markedness constraints are defined in terms of the Project and Pronounce relations. One constraint requires coda consonants, such as [h] in (2-19), to project a mora. This constraint is indifferent to whether the [h] is pronounced with its projected mora. Another markedness constraint requires that every mora be pronounced with some segment. This constraint is indifferent to whether the mora is pronounced with the segment that projects it. Though such mismatches between Project and Pronounce are possible — and (2-19) is an example — they are marked, violating a constraint called Reciprocity. This, in outline, is how opaque analyses are constructed in this theory.

What would it take to extend Turbidity theory to deal with the full range of opaque interactions? Very likely, it will require two coexistent phonological representations, the pronounced one and the projected one. These two representations are folded together in (2-19), but (2-19) is not representative of the full range of opaque interactions. An instructive example is Bedouin Arabic /gabr/ → [gabur], where /a/ is not raised in a derived open syllable. In Turbidity terms, this means that [gabur] must be represented with two coexistent syllabic parses, one where [b] projects as the coda of the syllable [gab] and one where it is pronounced as the onset of the syllable [bur]. Phenomena like Palestinian Arabic /fɑrˤm/ → [fɑrˤim] require extending the Project/Pronounce distinction to linear order relations among segments. Epenthetic [i] is pronounced as the successor to [rˤ] in the segmental string, but [m] is projected as [rˤ]'s successor. In short, there can be Project/Pronounce mismatches in all of the ways that phonological elements relate to one another. This means that there are two complete phonological representations, with two sets of markedness

constraints. Reciprocity maintains a check on divergence between the two representations, and violation of Reciprocity is the source of opacity.

Looked at in this way, Turbidity has much in common with those theories of opacity that posit a single additional level of representation besides underlying and surface structure. It also shares some of the limitations of these theories. We will examine those limitations in §2.3.4.2.

Segmental coalescence. Segmental coalescence may sound like a peculiar theory of opacity, but it plays such a role in much of the OT literature. It has been applied to one rather common form of counterbleeding opacity, in which a segment is observed to assimilate to another nearby segment that has deleted. The palatalization/syncope interaction in (2-5) is typical: /ħa:kim-i:n/ → [ħa:kʲmi:n]. Although this derivation is opaque under the assumption that palatalization is the result of assimilation, it can be analyzed as transparent under the assumption that palatalization and syncope are united into a single process of segmental coalescence. The underlying $/k_1i_2/$ sequence fuses into the single output segment $[k^j_{1,2}]$. There is no literal deletion, no failure of input-output correspondence, so Max is satisfied. The resulting segment is palatalized because $[k^j_{1,2}]$ is faithful to the color features of one of its underlying correspondents, $/i_2/$. Another alternative: when /i/ deletes, it leaves behind the feature specification [–back], which reassociates autosegmentally to the preceding /k/. In this case, although Max is violated, the feature-specific constraint Max(–back) is not. Analyses along these general lines can be found in Causley (1997), Gnanadesikan (1997, 2004), Lamontagne and Rice (1995), McCarthy and Prince (1995), and Pater (1996), among others.

These alternatives to opacity have their merits, but they also have their problems. A parochial concern is that palatalization in Bedouin Arabic is not limited to deleted /i/. Overt front vowels also cause palatalization, so palatalization must not be inextricably linked with deletion of the triggering segment, as both the coalescence and autosegmental analyses imply. A broader worry is that observed counterbleeding interactions are not conveniently limited to situations that can plausibly be regarded as coalescence. For example, Donegan and Stampe (1979: 153) point out that there is a counterbleeding interaction in English between intervocalic /t/-flapping and optional desyllabification of prevocalic liquids: /ʃætɹ̩ɪŋ/ $\rightarrow_{t\text{-flapping}}$ [ʃæɾɹ̩ɪŋ] $\rightarrow_{\text{optional}}$ [ʃæɾɹɪŋ] *shattering*. There is no way of reanalyzing this opaque interaction as a single coalescence process. Another example, this time from Kenstowicz and Kisseberth (1979: 292–294): in Tunica (Gulf, Louisiana), a sequence $/V_1$ʔɑ/ is altered by assimilating /ɑ/ to the color of $/V_1/$ and by deleting $/V_1/$ if it is unstressed. The result is a counterbleeding interaction in cases like /ˈhipu-ʔɑki/ → [ˈhipʔɔki] 'she dances', with /ɑ/ assimilating to the deleted vowel. The deletion + assimilation combination is unlikely to be reducible to a single process of coalescence for three reasons: (i) Deletion occurs independently of assimilation in cases like /ˈhɑrɑ-ʔuhki/ → [ˈhɑrʔuhki] 'he dances'; (ii) Assimilation occurs independently of deletion in cases like /ˈtʃu-ʔɑki/ → [ˈtʃuʔɔki] 'she takes'; and (iii) Coalescence of nonadjacent segments is probably unattested and very likely impossible (see §3.2.4.3).

Local constraint conjunction. As I noted in §2.3.3, counterfeeding opacity can be accommodated in classic OT if the theory of faithfulness constraints is sufficiently rich. (There is no comparable way of dealing with counterbleeding opacity.) It has been proposed that *local conjunction* of faithfulness constraints is the proper mechanism for incorporating this richer theory of faithfulness into Con (Ito and Mester 2003c, Kirchner 1996, Moreton and Smolensky 2002).

Local constraint conjunction is proposed by Smolensky (1995) as a theory of the internal structure of Con. Complex constraints are built by conjoining simpler constraints. (The simpler constraints may be irreducible, or they may themselves be the product of local conjunction.) The local conjunction of constraints A and B, $[A\&B]_{\delta}$, is defined as a constraint that is violated once for each instance of the domain δ in which both A and B are violated. Conjunction of markedness constraints supplies the most persuasive examples. Codas are marked by the constraint No-Coda and voiced obstruents are marked by the constraint No-Vcd-Obst. The local conjunction of these constraints within the domain of a segment, $[\text{No-Coda \& No-Vcd-Obst}]_{Seg}$, militates against the combination of these two marked properties, a voiced obstruent in coda position. In general, local conjunction of markedness constraints forbids the cooccurrence of marked structures in near proximity to one another.

In counterfeeding opacity, unfaithful mappings cannot occur in close proximity to one another. For instance, the counterfeeding interaction in Bedouin Arabic /gabr/ → [gabur] requires the local conjunction of Ident(low) and Dep in the domain of adjacent syllables: $[\text{Ident(low)\&Dep}]_{Adj\text{-}\sigma}$. By ranking $[\text{Ident(low)\&Dep}]_{Adj\text{-}\sigma}$ above the markedness constraint responsible for the open-syllable raising process (**a*CV in (2-20)), we ensure that opaque [gabur], which satisfies this constraint, is more harmonic than transparent *[gibur], which violates it.

(2-20) Counterfeeding opacity with local conjunction[13]

/gabr/	$[\text{Id(low)\&Dep}]_{Adj\text{-}\sigma}$	*Comp-Coda	**a*CV	Dep	Id(low)
→ gabur			1	1	
a. gibur	W_1		L	1	W_1
b. gabr		W	L	L	

This is a particularly elegant theory of counterfeeding opacity, but it cannot account for the full range of opacity phenomena, and it predicts a kind of pseudo-opacity that does not seem to exist (McCarthy 1999: 365–366, 2002a, 2003a, Padgett 2002). The reasons for both of these problems go right to the core of the local-conjunction theory: real counterfeeding opacity is a matter of forbidden process *interaction*, but local conjunction regulates process *proximity*. Interaction and proximity are two very different things, and it is a mistake to confound them.

The Bedouin Arabic example illustrates this mistake. The [gabur] example shows that raising is blocked when the epenthetic vowel *follows* the syllable with the potentially raised vowel. On the other hand, (2-21) shows that raising is not blocked when the (italicized) epenthetic vowel *precedes* the (boldface) raised vowel. The conjoined constraint $[\text{Ident(low)\&Dep}]_{Adj\text{-}\sigma}$ is unable to make this distinction, since it forbids raising and epenthesis in adjacent syllables, regardless of their linear order.

(2-21) Adjacent epenthesis and raising in Bedouin Arabic
/tˤarad ʁanam-ih/ [tˤa.ra.d*i*ʁ.ni.mih] 'he pursued his sheep'

This is not a mere technical glitch, to be solved with a more sophisticated theory of the domains of conjunction. Rather, it is a basic failure of principle. It is not an accident that raising is prohibited *before* an epenthetic vowel but allowed *after* one. When the epenthetic vowel follows, epenthesis interacts with raising, since following epenthesis puts the potentially raised vowel into an open syllable. When the epenthetic vowel precedes, however, epenthesis does not interact with raising, since preceding epenthesis has no effect on whether the potentially raised vowel is in an open syllable. Local conjunction uses proximity — the adjacent-syllables domain — as a proxy for interaction, and interaction is the real basis for opacity. Because phonological processes are usually locally conditioned, proximity is often successful as a proxy for opacity, but examples like this one decouple the effects of proximity and interaction, showing that interaction, not proximity, is what really matters.

Another way of grasping the problem with local conjunction's proximity = interaction equation is to look at the effects of locally conjoining faithfulness constraints in inappropriate domains. For example, the constraint $[\text{Ident(low)\&Dep}]_{Wd}$ — identical to Bedouin Arabic, except that the domain is larger — will block raising if a vowel has been epenthesized anywhere in the same word. If Bedouin Arabic were to have such a constraint, underlying /samiʕ-t-k/ 'I heard you (masculine singular)' would map to [samiʕt*a*k] instead of the expected [simiʕt*a*k]. No known language exhibits this sort of hyperopacity, in which counterfeeding behavior is extended from a local, interacting context to a distant, noninteracting context. Yet the local conjunction theory of counterfeeding opacity would seem to predict exactly this, since the domain of conjunction is stipulated independently of the conjoined constraint (Alderete 1997, Ito and Mester 2003a: 105ff.).

Similar problems arise when inappropriate constraint combinations are assembled by local conjunction. Imagine a language that is identical to Bedouin Arabic except that it also has final devoicing of obstruents. The conjoined constraint $[\text{IDENT(low)\&IDENT(voice)}]_{\text{Adj-}\sigma}$ could block raising whenever an adjoining syllable contains a devoiced obstruent: /katab/ → [**katap**]. This sort of hyperopacity is never attested — that is, we never find that one process blocks another if the two processes by their very nature cannot interact.

There have been efforts to impose restrictions on local conjunction to address some of these problems (Bakovic 1999, Fukazawa and Miglio 1998, Hewitt and Crowhurst 1996, Ito and Mester 2003a: 102ff., 2003c, Łubowicz 2002, 2006). Typically, these proposals rely on the shared formal properties of two constraints to determine whether they are conjoinable or, if conjoined, what their domain is. None of these proposals has been fully successful in addressing the problems described here and elsewhere (McCarthy 1999, 2002a, 2003a, Padgett 2002). The reason for this failure is not far to seek: counterfeeding opacity is a matter of forbidden process interaction, and process interaction is not something that can be determined solely by looking at the formal properties of faithfulness constraints.[14] Whether and under what conditions two processes will interact is something that depends on the circumstances that obtain in a particular language. We require a theory of opacity that is sensitive to these circumstances. Rule ordering is an example of such a theory, but others will be discussed here and in later chapters.

2.3.4.2 Analogues to serial derivations and rule ordering

Since rule-based phonology uses serial derivations to account for opacity, it is natural to ask whether derivations and the effects of rule-ordering can be reconstructed in OT, which is a theory without rules. A multi-step serial derivation can be obtained simply by assuming that the output of an OT grammar is not the surface form but instead is the input to another OT grammar. Actual implementations differ in whether or not the second grammar is the same as the first one. The approaches to be discussed are: single-grammar serial OT, which is known as harmonic serialism; multi-grammar serial OT, which is sometimes known as Stratal OT; and output-output faithfulness, often referred to as OO correspondence.

Harmonic serialism. In harmonic serialism, the output of an OT grammar is returned as the input to that same grammar (McCarthy 2000a, 2002b: 159–163, 2007a, Prince and Smolensky 2004: 6–7, 94–95). This process continues until 'convergence', when the output of a pass through the grammar is identical to the output of the previous pass. (Convergence in a finite number of passes is guaranteed for reasons discussed by Moreton (2003).)

Harmonic serialism in its simplest form turns out to be surprisingly ineffective in dealing with opacity. It is no better off than classic OT in dealing with counterbleeding opacity. In (2-16), we saw that classic OT stumbles on a case of counterbleeding opacity like /ħaːkim-iːn/ → [ħaːkʲmiːn] because there is no visible motive in surface structure for palatalization of the /k/. Harmonic serialism does no better. On the first pass through the grammar, there is nothing to prevent the transparent mapping /ħaːkim-iːn/ → *[ħaːkmiːn], just like (2-16). In general, wherever classic OT has a problem with counterbleeding opacity, harmonic serialism will too, since harmonic serialism is just classic OT, iterated.

Harmonic serialism actually does worse than classic OT on some kinds of counterfeeding opacity. Recall from (2-18) that classic OT can accommodate counterfeeding interactions by positing the right faithfulness constraints. In Bedouin Arabic, because /i/ deletes in the same environment where /a/ changes to [i], what is needed is a constraint that specifically militates against deleting /a/, MAX-A. But MAX-A is useless under the harmonic serialism regime. The first pass through the grammar, which is shown in tableau (2-22), maps /dafaʕ/ to [difaʕ]. The output of the first pass becomes the input to the second pass, shown in tableau (2-23), and [difaʕ] *qua* pass-two input is mapped to *[dfaʕ]. On the third pass, [dfaʕ] *qua* pass-three input maps to itself, and there is convergence — on the wrong output.

(2-22) Harmonic serialism: first pass through grammar

/dafaʕ/	MAX-A	*aCV	*iCV	ID(low)	MAX
→ difaʕ			1	1	
a. dfaʕ	W_1		L	L	W_1
b. dafaʕ		W_1	L	L	

(2-23) Harmonic serialism: second pass through grammar

[difaʕ]	Max-A	*aCV	*iCV	Id(low)	Max
→ dfaʕ					1
a. difaʕ			W_1		L
b. dafaʕ		W_1		W_1	L

The problem in (2-23) is this: with [difaʕ] as the input, Max-A no longer protects the vowel in the first syllable from deletion. When the second and subsequent passes through the grammar come around, information about the original input is no longer available to Eval. For this reason, counterfeeding opacity in general cannot be analyzed using harmonic serialism. (See Norton 2003: 247ff. for related discussion.)

These failures of harmonic serialism show that a single-grammar implementation of serial OT is of no value in analyzing opacity. We will see in §3.2.3, however, that harmonic serialism has some significant connections with OT-CC.

Multi-grammar serial OT. The principal thesis of the theory of Lexical Phonology is that the phonological system of a language consists of a series of separate modules, called levels or strata, each of which is an *SPE* grammar in its own right (see §2.2.6). Strata are usually associated with different morphological subsystems in the lexicon or with the difference between word-internal and phrasal phonology. There is an ordering among the strata, and the output of one stratum is the input to the next. The output of the last or postlexical stratum is the actual surface form.

It seems like a small step to go from assuming that strata are *SPE* grammars to assuming that they are OT grammars, and so this move has been advocated almost since the beginning of OT. Although implementational details differ, this idea of linking OT grammars serially is common to all of the approaches that go under names like LP/OT, Derivational OT, or Stratal OT. Throughout, I will use the name Stratal OT to refer to any theory that incorporates these basic assumptions.[15]

Stratal OT uses the ordering of strata to reproduce the effects of opaque ordering in rule-based phonology. If rule A precedes rule B in counterfeeding order in a rule-based analysis, then the Stratal OT reanalysis posits two strata. The grammar of the first stratum effects mappings equivalent to rule A, and the grammar of the second stratum effects mappings equivalent to rule B. The

output of the first stratum is the input to the second stratum just as the output of rule A is the input to rule B. Because of the assumed correlation between strata and morphological subsystems, the Stratal OT hypothesis about opacity is somewhat stronger than the rule-based hypothesis, which establishes no linkage between morphology and opacity.

Each stratum is an OT grammar, so within-stratum interactions are necessarily transparent just as they are in classic OT. The different strata are moreover *different* OT grammars from one another — that is, they are different permutations of the universal constraint set Con. This assumption is essential to Stratal OT's theory of opacity. Without it, Stratal OT would be another version of harmonic serialism, and we have already seen that harmonic serialism is a failed theory of opacity.

Stratal OT's central analytic strategy for opacity, then, is to isolate the opaquely interacting processes into different strata, with the ordering of the strata supplying the counterbleeding or counterfeeding order of the processes. For instance, the counterbleeding order between Bedouin Arabic palatalization and syncope in /ha:kim-i:n/ $\rightarrow_{\text{palatalization}}$ [ha:kʲim-i:n] $\rightarrow_{\text{syncope}}$ [ħa:kʲmi:n] shows that the stratum where palatalization occurs must be ordered before the stratum where syncope occurs. Because strata correlate with morphological subsystems or the lexical/postlexical distinction, it will sometimes be possible to use other evidence to determine exactly which strata are involved. Since syncope occurs in phrases as well as words (see (2-7)), it must occur in the postlexical stratum. For palatalization to precede syncope in counterfeeding order, palatalization must occur in some earlier, therefore lexical stratum. The OT grammar of the lexical stratum, shown in (2-24), takes the input /ha:kim-i:n/ and maps it to [ha:kʲimi:n], with palatalization but no syncope. The grammar of the postlexical stratum in (2-25) then takes [ha:kʲimi:n] as input and maps it to [ħa:kʲmi:n], with syncope. Observe that the two strata are inconsistent in how they rank Max and *iCV; they are, in every sense, different OT grammars.

(2-24) Lexical stratum

/ħa:kim-i:n/	Max	*ki	*iCV	Id(back)
→ ħa:kʲimi:n			1	1
a. ħa:kmi:n	W_1		L	L
b. ħa:kʲmi:n	W_1		L	1
c. ħa:kimi:n		W_1	1	L

(2-25) Postlexical stratum

/ħa:kʲimi:n/	*iCV	*ki	Max	Id(back)
→ ħa:kʲmi:n			1	
a. ħa:kmi:n			1	W_1
b. ħa:kʲimi:n	W_1		L	
c. ħa:kimi:n	W_1	W_1	L	W_1

There are two main problems with Stratal OT as a theory of opacity. First, Stratal OT is not powerful enough to deal with the full range of observed opaque interactions. Second, Stratal OT is also too powerful, since it massively overpredicts phonological systems that are never observed and seem impossible. There is, then, a two-way mismatch between the predictions of Stratal OT and the typology of known opaque interactions.

The argument that Stratal OT has insufficient power was foreshadowed at the end of §2.2.6. Like Stratal OT, rule-based Lexical Phonology allows for the possibility of between-stratum opaque orderings. But since each rule-based Lexical Phonology stratum is an *SPE* grammar, within-stratum opaque ordering is also possible. In general, the Lexical Phonology research program never sought to eliminate within-stratum rule ordering, including opaque ordering. In light of the extensive pre-Lexical Phonology literature arguing for the elimination of extrinsic ordering, this failure to pursue an obvious hypothesis might seem surprising, at least until one realizes the reason for it:[16] the hypothesis was self-evidently wrong. That is, research on rule-based Lexical Phonology never progressed in the direction of eliminating within-stratum opaque ordering because there was no shortage of Lexical Phonology analyses that crucially relied on such ordering, such as Kiparsky's (1984) analysis of Icelandic or Kiparsky's (1985) analyses of Catalan and Russian.

In Catalan, for example, there is a counterbleeding relationship between nasal place assimilation and final cluster simplification. According to Kiparsky, cluster simplification must be assigned to the lexical stratum because clusters cannot be rescued by postlexical resyllabification before vowel-initial words: *pont antic* 'old bridge' is pronounced as [ˌpɔ.nən.ˈtik] and not *[ˌpɔn.tən.ˈtik]. Since nasal place assimilation precedes cluster simplification, as shown in (2–26), nasal place assimilation must also apply in the lexical stratum. The result in this case, as in so many other Lexical Phonology analyses, is a within-stratum opaque (counterbleeding) order.

(2-26) Counterbleeding order in Catalan (Kiparsky 1985:96–97)

Underlying	/bɛn-k/
Place assimilation	[bɛŋk]
Cluster simplification	[bɛŋ]
Surface	[ˈbɛŋ]
	'I sell'

Bedouin Arabic supplies another example of the insufficiency of between-stratum ordering as a theory of opaque rule ordering. (See §4.3.3 for details.) In a rule-based analysis, deletion of high vowels must precede raising of low vowels in counterfeeding order: /dafaʕ/ $\rightarrow_{deletion}$ *DNA*[17] $\rightarrow_{raising}$ [difaʕ] (see (2-4)). We know from (2-7) that deletion is a process of the phrasal phonology, so it must occur as late as the postlexical stratum. Raising, on the other hand, is not a phrasal process — it only applies within words and never when its open-syllable context arises by resyllabification across word juncture. Therefore, raising cannot occur later than the last lexical stratum. Since raising is lexical and deletion is postlexical, these processes are intrinsically ordered by virtue of their stratal assignments, and so raising *must* precede deletion. But this is exactly the wrong conclusion, since it puts them in feeding order rather than counterfeeding order. Because raising is lexical, the lexical stratum maps /dafaʕ/ to [difaʕ], and because deletion is postlexical, the postlexical stratum goes on to map [difaʕ] to *[dfaʕ] (just as it maps /difiʕ/ 'was pushed' to [dfiʕ]).

There is more to be said about this example. As I showed in (2-18), classic OT can analyze this counterfeeding interaction if it has a constraint Max-A. This constraint prevents deletion of any underlying /a/, even if its surface realization is something other than [a]. Max-A is of no help in the stratal account, however. The problem is that the lexical stratum output [difaʕ] is the postlexical stratum input, and so the postlexical phonology sees an input [i] in the first syllable of this word. Max-A does not protect input [i]s from deletion. In Stratal OT, faithfulness constraints are local to each stratum: they require identity between that stratum's input and its output, and they have no way of accessing the original underlying representation /dafaʕ/. The information that the first vowel of [difaʕ] is an erstwhile /a/ has been lost irretrievably by the time the postlexical stratum comes along, and so neither Max-A nor any other constraint can account for the counterfeeding interaction between these two processes whose stratal assignments place them in feeding order. With respect to this example, then, Stratal OT is actually worse off than classic OT.

The Catalan and Arabic examples reveal some general properties of theories that seek to reduce opaque interactions to between-stratum orderings. If rule A precedes rule B in counterbleeding order, then A must apply on some stratum that is earlier than the stratum where B first applies. A may continue to apply on later strata, but A's earliest application must precede B's earliest application. If rule A precedes rule B in counterfeeding order, then A must apply on some stratum that is earlier than the stratum where B first applies, and A must not apply on B's earliest stratum or any subsequent stratum. These entailments of Stratal OT tell us what situations would constitute *prima facie* counterexamples to this theory of opacity, such as counterbleeding order with B in the earliest stratum or counterfeeding order with A in the last stratum. A specific prediction: postlexical processes like syncope in Bedouin Arabic are never opaque. See §4.3 for various demonstrations that this process is indeed opaque.

From the examples discussed, it appears that Stratal OT's premises are insufficient to account for the full range of observed opaque interactions (see also Noyer 1997: 515, Paradis 1997: 542, Roca 1997b: 14ff., Rubach 1997: 578 for similar remarks). The literature in support of Stratal OT and its variants has mostly focused on exhibiting between-stratum opaque interactions and arguing against other approaches to opacity in OT, such as sympathy theory (§2.3.4.3). I am not aware of comparable work arguing that Stratal OT is sufficient to account for the full range of observed opaque interactions. The evidence described here and in the Lexical Phonology literature challenges this claim.

Stratal OT is also an overly powerful theory because it imposes no limits on differences among strata within a single language. There is a profound but mostly unacknowledged difference between rule-based Lexical Phonology and Stratal OT on exactly this point. Each Lexical Phonology stratum is an *SPE* grammar and each Stratal OT stratum is an OT grammar. This seeming parallelism is misleading, however, because the literature on rule-based Lexical Phonology was highly attentive to the problem of constraining between-stratum differences. There are serious and well-argued (though not uncontroversial) proposals about how to do this. The earliest proposals took the form of principles for separating lexical and postlexical processes (e.g., Kaisse and Hargus 1993a: 16–17, Kiparsky 1983, Mohanan 1982): lexical rules are structure-preserving (i.e., neutralizing or nonallophonic); lexical rules are word-bounded; lexical rules apply only in derived environments; lexical rules apply only to the lexical categories noun, verb, and adjective; only lexical rules may have exceptions; only lexical rules are sensitive to word-internal morphological structure; and lexical rules are categorical, never gradient. This body of work culminated in the Strong Domain Hypothesis (Borowsky 1986, Kiparsky 1984, Myers 1991b, Selkirk 1982b): all strata, lexical and postlexical, share a single *SPE*-type grammar. The observed differences between strata are obtained from a combination of universal principles like structure preservation, which can prevent some rules from applying in lexical strata, and language-particular stipulations about when certain rules stop applying. An approach like this is clearly far more restrictive than the original Lexical Phonology thesis that each stratum is a separate *SPE* grammar.

This restrictive version of Lexical Phonology cannot be reconstructed in Stratal OT, however. Structure preservation, for example, is the cornerstone of the Strong Domain Hypothesis, but there is no hope of developing an analogue to structure preservation in Stratal OT. The principle of structure preservation says that rule application in lexical strata cannot create segments or structures that are not already present in underlying representations. In other words, the well-formedness conditions on underlying representations persist as conditions on rule application throughout the lexical strata, although they may be relaxed or turned off in the postlexical stratum.

Structure preservation has no OT analogue for two reasons:

> First, the hypothesis that grammars differ only in constraint ranking entails that there can be no language-particular conditions on underlying representation (McCarthy 2002b: 70–71, Prince and Smolensky 2004). This requirement is called richness of the base (ROTB) (see also §3.5.2). Under ROTB, *the grammar itself, unaided by restrictions on its inputs*, is responsible for observed phonotactic patterns. Since there are no restrictions on inputs, it would make no sense to speak of such restrictions persisting in their effects through the lexical strata.
>
> Second, OT offers no way of reconstructing rule-based Lexical Phonology's notion that some lexical constraints are turned off in later strata. The naïve supposition is that turning-off effects can be simulated by demoting markedness constraints or promoting faithfulness constraints. In reality, though, OT offers no simple equivalence between demotion or promotion and deactivation. Even low-ranking markedness constraints may be active in situations where the faithfulness constraints ranked above them are not relevant. Thus, the specific effects of markedness demotion or faithfulness promotion cannot be predicted without meticulous examination of the entire constraint hierarchy and array of inputs. Prince and Smolensky (2004: 27ff.) emphasize this point for faithfulness constraints; reduplicative emergence of the unmarked illustrates the same point for markedness constraints (Alderete *et al.* 1999, McCarthy and Prince 1994). Known conditions of literal deactivation of a constraint, such as Panini's Theorem (Prince and Smolensky 2004: 97–99), have such specific conditions that they are of little value in characterizing permitted differences between strata.

It follows, then, that the restrictive theory of differences between strata that was developed in rule-based Lexical Phonology does not and presumably cannot inform our understanding of such differences in Stratal OT. This problem is not unknown to proponents of Stratal OT, and they have attempted to develop simple principles for relating the constraint hierarchies of different strata within a language. An example: Kiparsky (1997: 17) proposes that the ranking of markedness constraints is constant across all of the strata of a language, so between-stratum differences are limited to promotions and demotions of faithfulness constraints. Another example: Koontz-Garboden (2003), citing a personal communication from Kiparsky, proposes that between-stratum reranking is limited to promoting constraints to undominated status in later strata (that is, stratum n+1 is identical to stratum n except that some lower-ranking constraint(s) in n are undominated in n+1). It is not hard to find counterexamples to these hypotheses in the Stratal OT literature. For example, Ito and Mester (2001: 274–276) argue that the lexical and postlexical strata in German differ in markedness ranking, contrary to the first hypothesis. The second hypothesis is inconsistent with Kiparsky's (2003) analysis of syncope in colloquial Arabic. In that analysis, syncope is the result of satisfying a constraint against light syllables. This constraint is promoted to a higher rank in the word stratum than in the earlier, stem stratum. But this promotion cannot be to undominated status; the constraint against light syllables must be dominated since the language has some light syllables that escape the effects of syncope. If it were undominated, then the language could have no light syllables whatsoever.

In summary, Stratal OT has not and probably cannot recapture Lexical Phonology's restrictive theory of between-stratum differences, nor does it yet have a workable substitute. Absent such restrictions, Stratal OT allows the strata of a single language to differ by as much as one language differs from another. A stratum is just a ranking of Con, with no obligations to the rankings of Con in other strata of the same language. From the perspective of language typology and learnability, this is an unwelcome conclusion.

Output-output faithfulness. The theory of output-output correspondence posits faithfulness relations among morphologically related output forms (Benua 1997, Kenstowicz 1996a, Pater 2000, and many others). Information flows via faithfulness constraints from an output form called the 'base' to other forms derived from it by affixation. OO correspondence can be applied to phonological opacity, as in Kager (1999b). If the base transparently undergoes or fails to undergo the potentially opaque process, then OO faithfulness constraints can transmit this information by compelling related forms to resemble the base. For example, the opaquely palatalized velar in Bedouin Arabic /ħa:kim-i:n/ → [ħa:kʲmi:n] could be explained with reference to the unaffixed singular base form [ħa:kʲim], where palatalization is transparent.

This is accomplished formally by deploying the output-output faithfulness constraint OO-Ident(back), ranking it higher than its input-output counterpart IO-Ident(back) (which is referred to as just Ident(back) in tableau (2-16)). In this way, the phonologically unremarkable velar palatalization in [ħa:kʲim] is transmitted to the rest of the paradigm, even to forms where the triggering front vowel is absent from surface structure.

OO faithfulness does not suffice as a theory of opacity, however. There are three main arguments against it (also see Benua 1997, Booij 1996, 1997, Ito and Mester 1997b, Karvonen and Sherman [Ussishkin] 1998, McCarthy 1999: 385–387, Noyer 1997, Paradis 1997, Rubach 1997).

First, it is impossible to use OO faithfulness as a comprehensive theory of opacity and also have a principled theory of what can be the base of an OO correspondence relation. The [ħa:kʲmi:n] example is attractive because the morphologically basic form is the one where the process is transparent and the forms derived from it are the ones where the process is opaque. In the Bedouin Arabic counterfeeding case /gabr/ → [gabur], however, there is no word that is morphologically more basic than [gabur] 'grave'. The only paradigm members where the lack of raising can be explained transparently are the derived forms, such as [gabri] 'my grave'. Clearly, it is unreasonable to insist that [gabur] 'grave' is derived from [gabri] 'my grave', but that is exactly what would be required to account for [gabur]'s unraised vowel using OO faithfulness.

Another argument against OO faithfulness as a theory of opacity is the existence of cases where the (non-)application of a process is transparent in *no* member of the paradigm. The analysis of Tiberian Hebrew epenthesis in McCarthy (1999) is an example. Alternations like those in (2-27) show that surface [ˈpɛlɛ] is derived from an underlying form with a final glottal stop, /pɛlʔ/. In a rule-based analysis (Malone 1993: 59–60, 93–94, Prince 1975: 37ff.), this mapping is the result of three processes applied in the opaque order shown in (2-28).[18] Epenthesis renders stress opaque, and deletion of [ʔ] renders epenthesis opaque.

(2-27) Alternations of underlying /pɛlʔ/

/pɛlʔ/	ˈpɛlɛ	'a wonder' (Exodus 15, verse 11)
/pɛlʔ-aka:/	pilʔăˈxa:	'your wonder' (Psalms 89, verse 6)
/pɛlʔ-i:m/	pəla:ˈʔi:m	'wonders' (Lamentations 1, verse 9)

(2-28) Tiberian Hebrew /pɛlʔ/ → [ˈpɛlɛ] derivation

Underlying	/pɛlʔ/
Stress final closed syllable	ˈpɛlʔ
Epenthesis in final cluster	ˈpɛlɛʔ
Deletion of final [ʔ]	ˈpɛlɛ
Surface	[ˈpɛlɛ]

To analyze opaque stress and epenthesis using OO faithfulness, at a minimum we would need to find paradigm members where stress on [pɛ] and epenthesis between [l] and [ʔ] are occurring transparently. There are none. The rest of the paradigm consists of words with vowel-initial suffixes. Because of these suffixes, stress is never retracted as far as [pɛ] and epenthesis is unnecessary. Therefore, neither of these opaque phenomena can be obtained with OO correspondence constraints.

The same problem for OO faithfulness — transparency nowhere in the paradigm — arises whenever an underlying phonological contrast undergoes absolute neutralization. For example, the underlying pharyngeal /ʕ/ in Maltese appears to condition a number of phonological processes, though it is always deleted at the surface (Borg 1997, Brame 1972). One such process lowers vowels next to pharyngeal consonants: /nimsiħ/ → [nimsaħ] 'I wipe'. This process is conditioned opaquely by the deleted /ʕ/: /nismiʕ/ → [nisma] 'I hear'. Nowhere in the paradigm of /smiʕ/ or, indeed, any other word of standard Maltese is the /ʕ/ preserved on the surface, to condition lowering transparently.

The Hebrew and Maltese critiques of OO faithfulness as a theory of opacity apply with equal force to approaches based on paradigm uniformity (see Downing, Hall, and Raffelsiefen (eds), (2005)). Paradigm uniformity allows information to flow in any direction among paradigm members, so a morphologically complex form can affect a simple form. This greater freedom is useless, however, in analyzing opacity when no member of the paradigm meets the transparency requirement.

A final argument against OO faithfulness as a theory of opacity is the existence of cases where OO faithfulness overpredicts opaque behavior. In Levantine Arabic (see §4.2), stress and epenthesis interact opaquely, leading to surface contrasts like [ˈkatabit] 'she wrote' (from /katab-it/) vs. [kaˈtab*i*t] 'I wrote' (from /katab-t/). Stress is assigned transparently in /katab-it/ → [ˈkatabit], but stress is assigned opaquely in /katab-t/ → [kaˈtab*i*t], as if the epenthetic vowel were not present. To account for the opaque stress of [kaˈtab*i*t] 'I wrote' in OO faithfulness terms, we would need to explain why this form is taking its cues, stress-wise, from paradigm members like [kaˈtabna] 'we wrote' and not from [ˈkatabit] 'she wrote' or [ˈkatab] 'he wrote'. Furthermore, we would need to explain why a high-ranking OO faithfulness or paradigm uniformity constraint affects only [kaˈtab*i*t], the form that just happens to contain an epenthetic vowel. Any OO faithfulness constraint that would affect [kaˈtab*i*t] would surely resist all stress alternations throughout the paradigm, so we would expect consistent stress on the second syllable: *[kaˈtabit] for 'she wrote', *[kaˈtab] for 'he wrote', and so on. An OO faithfulness analysis of these facts seems quite hopeless.

OO faithfulness's inadequacy as a theory of opacity is not entirely unexpected. OO faithfulness is a reasonable theory of phonological similarity among morphologically related forms, but this is a far cry from opacity's hidden generalizations.

2.3.4.3 Analogues to intermediate derivational forms

Sympathy (McCarthy 1999, 2003c), targeted constraints (Wilson 2000), enriched inputs (Sprouse 1997, 1998), and comparative markedness (McCarthy 2003a, 2003d) are four theories of opacity in OT that share a commitment to using a third form, neither input nor output, in candidate evaluation. Since sympathy theory has been examined more extensively than the other approaches, the discussion here will focus on it exclusively.[19]

In sympathy theory, the third form is called the sympathetic candidate. The sympathetic candidate is just that, a candidate, so it is a kind of output form, though different from the actual output. The sympathetic form differs from the actual output by virtue of satisfying some faithfulness constraint that the actual output violates. This faithfulness constraint is called the selector. Apart from obeying the selector, the sympathetic candidate is as harmonic as possible; it is, in short, the most harmonic candidate among those that obey the selector. For example, in the Bedouin Arabic palatalization/syncope interaction (2-5), the selector constraint is MAX, so the sympathetic candidate does not have syncope, though the actual output does. But because the sympathetic candidate is maximally harmonic in all other respects, it shows the effects of all of the other (transparent) phonology of the language. Therefore, the sympathetic candidate from input /ħa:kim-i:n/ is [ħa:kʲimi:n], without syncope but with palatalization, since palatalization of velars is required before front vowels.

The sympathetic candidate influences the choice of the actual output form by way of sympathy constraints. Sympathy constraints look like faithfulness constraints, but they evaluate resemblance to the sympathetic candidate rather than resemblance to the input. With the right ranking, as shown in (2-29), the sympathy constraint favors opaque [ħa:kʲmi:n], whose palatalization matches sympathetic [ħa:kʲimi:n], over the transparent form [ħa:kmi:n], which has no palatalization. Two candidates in (2-29) obey the selector constraint MAX, (b) and (c). Of these, (b) is more harmonic by virtue of satisfying **ki*, so it is the sympathetic candidate. (To avoid circularity, the sympathy constraint itself must be ignored in determining the sympathetic candidate.) The sympathy constraint IDENT(back)$_{Sym}$ is satisfied by candidates that match the palatalization in the sympathetic candidate, thereby ruling out (a). Recall from (2-16) that (a) harmonically bounds the intended winner in classic OT. The sympathy constraint breaks this harmonic bounding.

(2-29) Counterbleeding opacity with sympathy

/ħa:kim-i:n/	Id(back)$_{Sym}$ (sympathy)	*iCV	*ki	Max (selector)	Id(back)
→ ħa:kʲmi:n				1	1
a. ħa:kmi:n	W$_1$			1	L
b. ħa:kʲimi:n (sympathetic cand.)		W$_1$		L	1
c. ħa:kimi:n	W$_1$	W$_1$	W$_1$	L	L

The sympathetic candidate [ħa:kʲimi:n] is identical with the intermediate stage of the serial derivation (2-5). *This is no accident. The changes that sympathy* theory requires in classic OT are not so different from the changes that are required in a derivational approach like Stratal OT. The selection of the sympathetic candidate requires a separate harmonic evaluation in which a single faithfulness constraint, the selector, is promoted to undominated status. Except for the more limited reranking possibilities, this is not unlike the grammars of different strata in Stratal OT. Furthermore, there is a fundamental asymmetry between the sympathetic candidate and real output candidates: the sympathetic candidate influences the choice of the actual output, but the actual output is not allowed to influence the choice of the sympathetic candidate. This asymmetry is a necessary property of serial derivations: the early stages of the derivation influence the later stages, and not vice-versa.

Various objections have been raised against sympathy theory (Bye 2001, Idsardi 1997, Ito and Mester 2001, Kiparsky 2000, McMahon 2000), but the biggest problem may be the analysis of multiple interacting opaque processes. The most famous example of multiple interaction is Yawelmani Yokuts. (References on this language include Archangeli 1985, Archangeli and Suzuki 1996, 1997, Cole and Kisseberth 1995, Dell 1973, Goldsmith 1993a, Hockett 1973, Kenstowicz and Kisseberth 1977, 1979, Kisseberth 1969, Kuroda 1967, Lakoff 1993, Newman 1944 (the original source), Noske 1984, Prince 1987, Steriade 1986, Wheeler and Touretzky 1993, Zoll 1993.)

Yawelmani has three processes that interact opaquely:

a) *Height-stratified rounding harmony:* a suffix vowel takes on the rounding of the preceding vowel if they agree in height. E.g., the high-voweled nonfuture suffix /-hin/ alternates as follows: [dubh**u**n] 'lead by the hand' vs. [bokˀ**hi**n] 'find'; cf. [xathin] 'eat', [xilhin] 'tangle'. In contrast, the nonhigh-voweled dubitative suffix /-al/ alternates like this: [kˀoʔ**o**l] 'throw' vs. [hud**a**l] 'recognize'; cf. [maxal] 'procure', [gijˀal] 'touch'.

b) *Long-vowel lowering:* underlying long high vowels become mid. E.g., /ʔili:-hin/ → [ʔil**e:**hin] 'fan', /cˀuju:-hin/ → [cˀuj**o:**hun] 'urinate'.

c) *Closed syllable shortening:* long vowels are shortened in medial and final closed syllables. E.g., /ṣa:p-hin/ → [ṣ**a**phin] 'burn', /pana:-al/ → [pan**a**l] 'arrive'.

In a rule-based analysis or its Stratal OT counterpart (Kiparsky 2001), these three processes must apply in the order shown in (2-30). (For a different analysis of Yawelmani, see §3.3.5.) No other order of application will do. If the order of rounding harmony and lowering were reversed, then the high suffix vowel would not harmonize with an originally high root vowel that has been lowered because of its length: /ʔu:t̪-hin/ $\rightarrow_{\text{lowering}}$ [ʔ**o:**t̪-hin] $\rightarrow_{\text{harmony}}$ *DNA* $\rightarrow_{\text{shortening}}$ *[ʔ**o**t̪hin]. And if the order of lowering and closed syllable shortening were reversed, then the derived short vowel would fail to lower: /ʔu:t̪-hin/ $\rightarrow_{\text{harmony}}$ [ʔu:t̪h**u**n] $\rightarrow_{\text{shortening}}$ [ʔ**u**t̪hun] $\rightarrow_{\text{lowering}}$ *DNA* $\rightarrow_{\text{surface}}$ *[ʔut̪hun].

(2-30) Rule-based or Stratal OT derivation for Yawelmani

Underlying	/ʔu:t̪-hin/	'steal'
Rounding harmony	ʔu:t̪hun	
Long-vowel lowering	ʔo:t̪hun	
Closed syllable shortening	ʔot̪hun	
Surface	[ʔot̪hun]	

Unsurprisingly, Yawelmani's multiple opaque interactions cannot be analyzed with a single selector constraint choosing a single sympathetic candidate. Two selector constraints, two sympathetic candidates, and two sympathy constraints are required. One of the selectors is the antishortening constraint Max-μ, and from the input /ʔu:t̪-hin/ it favors the sympathetic candidate [ʔo:t̪hin]. The other selector is Ident(high), and it favors the sympathetic candidate [ʔut̪hun]. Each sympathetic candidate has its own sympathy constraint. The role of the sympathetic candidate [ʔo:t̪hin] is to determine the height of the actual output [ʔot̪hun]'s root vowel. It does this by way of a sympathy constraint called Ident(high)$_{Max\text{-}\mu}$. (The Max-μ subscript is there to index this sympathy constraint to the selector of its sympathetic candidate.) The other sympathetic candidate, [ʔut̪hun], determines the rounding of [ʔot̪hun]'s suffix vowel. It does this by way of the sympathy constraint Ident(round)$_{Ident(high)}$. In sum, the output's vowel height is taken from the sympathetic candidate selected by

Max-μ, whereas the output's rounding comes from the sympathetic candidate selected by Ident(high).

Once sympathy theory is provided with these additional analytic resources to handle multiple opacity, however, it is in serious danger of overgeneration. Kiparsky (2001) presents a simple but striking example. Assume that Con has only the markedness constraint No-Coda and consonant- and vowel-specific versions of the faithfulness constraints Dep and Max. As shown in (2-31), Dep-V and Max-C can each act as a selector. Given the input /pam/ and a ranking where No-Coda dominates both Dep-V and Max-C, Dep-V selects the sympathetic candidate [pa], which deals with the potential coda by deletion, and Max-C selects the sympathetic candidate [pamə], which deals with the potential coda by epenthesis. The sympathy constraints favor outputs that resemble these two sympathetic candidates in specific ways. The sympathy constraint Dep-$C_{\text{Dep-V}}$ favors any candidate that has no consonants that are not present in the Dep-V-selected sympathetic candidate [pa]. This means that Dep-$C_{\text{Dep-V}}$ favors forms that replicate [pa]'s consonant deletion. Similarly, the sympathy constraint Max-$V_{\text{Max-C}}$ favors any candidate that has all of the vowels that are present in the Max-C-selected sympathetic candidate [pamə]. This means that Max-$V_{\text{Max-C}}$ favors forms that replicate [pamə]'s vowel epenthesis. The net result is that the winner is [paə], a form that reproduces both [pa]'s consonant deletion and [pamə]'s vowel epenthesis.

(2-31) An unwelcome result of sympathy

/pam/	Dep-$C_{\text{Dep-V}}$ (sympathy)	Max-$V_{\text{Max-C}}$ (sympathy)	No-Coda	Dep-V (selector)	Max-C (selector)
→ paə				1	1
a. pam	W_1	W_1	W_1	L	L
b. pamə (sympathetic via Max-C)	W_1			1	L
c. pa (sympathetic via Dep-V)		W_1		L	1

The problem with (2-31) is that this sort of opaque interaction is unattested and no doubt impossible. The sympathetic candidates reflect two different ways of satisfying No-Coda, epenthesis and deletion. The sympathy constraints force the winner to reproduce the effects of *both* ways of satisfying No-Coda, both

epenthesizing and deleting when either one alone would be enough. Obviously, additional constraints could be introduced to rule out [paə], but this sort of local fix misses the broader point. This is simply not an attested opaque interaction; it is gratuitous unfaithfulness, a kind of hyperopacity. It would seem, then, that giving sympathy the power to deal with multiple opacity also gives it the power to produce such unlikely results as (2-31).

A final remark about sympathy. Bye (2001, 2003), Jun (1999), and Odden (1997) introduce variations on sympathy theory that make the sympathetic form part of the candidate that is evaluated. One way of implementing this idea is to assume that a candidate is not a single form but rather an ordered pair: (*sympathetic-form*, *output-form*). In Bedouin Arabic, for example, the winning candidate would be ([ħa:kʲimi:n], [ħa:kʲmi:n]), and it competes against alternatives like those listed in (2-32). There are various ways of constructing a system of constraints for evaluating candidates like these; the comments at the right in (2-32) give a sense of what the constraints will need to be sensitive to. As we will see in §3, candidate chains are a somewhat similar idea.

(2-32) Candidates as (*sympathetic-form*, *output-form*)

([ħa:kʲimi:n], [ħa:kʲmi:n])	Winner.
([ħa:kʲimi:n], [ħa:kmi:n])	Palatalization mismatch.
([ħa:kimi:n], [ħa:kmi:n])	No palatalization in the sympathetic form.
([ħa:kmi:n], [ħa:kmi:n])	Syncope in the sympathetic form.

2.3.4.4 *Opacity as a mechanism for preserving contrasts*

Donegan and Stampe (1979), Kaye (1974, 1975), Kisseberth (1976), and Gussmann (1976) propose that opaque rule ordering has a functional explanation. The general idea is that opacity preserves phonemic contrasts, avoiding neutralizations that would occur if the rules applied in transparent order.

Kaye (1974) looks at counterbleeding orders like the one in (2-33), which comes from Ojibwa (Algonquian, US and Canada). This is a counterbleeding order because, if the rules were applied in the opposite order, cluster simplification would deprive place assimilation of an opportunity to apply, yielding *[takoʃʃin] instead (cf. Catalan). Kaye observes that the opaque order makes sense functionally: to a listener hearing the surface form, '… *it is immediately* apparent that the underlying representation ends in *k*, given the rules cited above and the fact that *ŋ* is not part of the inventory of underlying segments. … The only possible source of *ŋ* is as a result of the assimilation of a nasal to a following velar stop' (Kaye 1974: 144). Kaye uses the term 'recoverability' to describe this functional motivation for opacity.

(2-33) Counterbleeding order in Ojibwa (Kaye 1974: 140)

Underlying	/takossin-k/	
Place assimilation	takossiŋk	
Cluster simplification	takossiŋ	
Surface	[takoʃʃiŋ]	'(if) he arrives'

Donegan and Stampe (1979: 145–151) make a somewhat similar point about counterfeeding order. They see phonology as the result of conflict between phonetic (articulatory) and phonological (perceptual) aims. Transparent interaction of processes is phonetically motivated, since it presumably maximizes articulatory ease. Opaque interactions of the non-surface-true variety are phonologically motivated, in their sense, because opacity 'bring[s] speech closer to its phonological intentions' (p. 147). As an example, Donegan and Stampe cite nasal deletion and intervocalic flapping in English *plant it*. For some speakers, they interact opaquely, as shown in (2-34). (Compare [plæ̃t ɪt] with [plæ̃ɾ̃ ɪ̃t], which other speakers produce from transparent interaction of the same processes.) In Donegan and Stampe's view, this and other instances of counterfeeding order 'prevent the merger of phonologically distinct representations' (p. 147), such as *plant it* and *plan it*. The desire to avoid merger must be weighed against the cost of opaque [plæ̃t ɪt]'s greater articulatory difficulty in comparison with transparent [plæ̃ɾ̃ ɪ̃t].

(2-34) Counterfeeding order in English *plant it*

Underlying	/plænt ɪt/
[nasal] assimilation	plæ̃nt ɪt
[t] flapping	*Inapplicable because [t] is not intervocalic*
Nasal deletion	plæ̃t ɪt
Surface	[plæ̃t ɪt]

Łubowicz (2003) develops an Optimality-Theoretic system, called PC theory (for 'preserve contrast'), in which these ideas about opacity's functional motivation are given a formal basis. With Flemming (1995), Padgett (2003), and others, she assumes that the objects of phonological evaluation are systems of contrasts, which she calls scenarios, rather than individual forms. (E.g., a scenario for German /bund/ 'federation' might include all of its logically possible minimal pairs, including /bunt/.) This move allows Con to include constraints against neutralization, which can favor opaque interactions precisely when they help to preserve a contrast that would otherwise be lost.

Łubowicz applies PC theory to, among other things, counterfeeding opacity. Take an example like Bedouin Arabic /gabr/ → [gabur] (see (2-20)). Because the output contains an unraised vowel in an open syllable, it offers a hint that the openness of the syllable is not original. PC theory expresses this intuition formally by introducing constraints on scenarios, among which is one called PC_{IN}(V/Ø). This constraint is violated by any output scenario that neutralizes a contrast between a vowel and zero that obtains in the input scenario. Importantly, PC_{IN}(V/Ø) does not say *how* the contrast is to be preserved; it can be preserved as-is, or it can be transferred to some other segment, depending on interaction with other constraints. Thus, if /pat/ and /pati/ both map to [pati], PC_{IN}(V/Ø) is violated, but it is not violated if /pat/ maps to [pati] and /pati/ maps to [padi].

In the Arabic case, PC_{IN}(V/Ø) evaluates scenarios like those in (2-35). (Underlying /gabur/ is hypothetical; the scenarios deal with possible rather than actual words.) To block raising and thereby favor the opaque scenario, PC_{IN}(V/Ø) must crucially dominate the markedness constraint that favors raising, **a*CV. The effect of the ranking [PC_{IN}(V/Ø) >> **a*CV] in (2-36) is that the underlying contrast between a vowel and zero, which epenthesis threatens to neutralize, is transferred to the quality of the vowel of the preceding syllable. This brief analysis glosses over some important issues, but it is sufficient to get a sense of how PC theory works and how it can account for opacity.

(2-35) Arabic scenarios in PC theory

a. Transparent
/gabur/ → [gibur]
/gabr/ → [gibur]

b. Opaque
/gabur/ → [gibur]
/gabr/ → [gabur]

(2-36) Counterfeeding opacity in PC theory

	/gabur/ /gabr/	PC_{IN}(V/Ø)	*Comp-Coda	**a*CV	Dep	Id(low)
→	/gabur/ → [gibur] /gabr/ → [gabur]			1	1	1
a.	/gabur/ → [gibur] /gabr/ → [gibur]	W_1		L	1	W_2
b.	/gabur/ → [gibur] /gabr/ → [gabr]		W_1	L	L	1

Łubowicz does not discuss counterbleeding opacity, but PC theory also seems applicable to cases like Kaye's Ojibwa example. Two of the relevant scenarios are given in (2-37). The transparent scenario is clearly less marked and more faithful, yet the opaque one wins. For this to happen, PC_{IN}(C/Ø) must be ranked above the markedness and faithfulness constraints that favor the transparent scenario. This ranking argument is shown in (2-38). The idea is that the contrast between /k/ and Ø is preserved by being transferred onto the preceding consonant.

(2-37) Ojibwa scenarios in PC theory

a. Transparent

/...in/ → [...in]
/...ink/ → [...in]

b. Opaque

/...in/ → [...in]
/...ink/ → [...iŋ]

(2-38) Counterbleeding opacity in PC theory

	/...in/ /...ink/	PC_{IN}(C/Ø)	*COMP-CODA	*ŋ	ID(coronal)
→	/...in/ → [...in] /...ink/ → [...iŋ]			1	1
a.	/...in/ → [...in] /...ink/ → [...in]	W_1		L	L
b.	/...in/ → [...in] /...ink/ → [...ink]		W_1	L	L

PC theory is by far the most original theory of opacity among those we have seen, since it offers a radical restatement of the entire rationale for opaque interactions. But it is not without its problems, the most serious of which can be illustrated with the Ojibwa example in (2-38). The constraint PC_{IN}(C/Ø) chooses the opaque winner over its transparent competitor in (a). That may seem unexceptionable, but in a way the PC constraint works *too* well. The victory of the opaque scenario in (2-38) is unrelated to the fact that Ojibwa independently has a process of nasal place assimilation (e.g., [takoʃʃiŋkipan] '(if) he arrived then'). In (2-38), the winner manages to beat (a) without reference to any markedness constraint that favors [ŋk] over [nk]. The analysis in (2-38) therefore makes no connection between how contrast is preserved and the independent existence of a nasal place assimilation process in the language.

This means that, without significant modifications, PC theory allows contrasts to be preserved in ways that are fundamentally unnatural, since they do not depend on markedness constraints for their motivation (Łubowicz 2003: 148–153). For example, the contrast between /k/ and Ø in Ojibwa could in principle be preserved by rounding the preceding vowel: /...in/ → [...in], /...ink/ → [...yn]. This outcome is made possible by ranking PC_{IN}(C/Ø) above **y* and IDENT(round), even though neither Ojibwa nor any other language is likely to have a process that transparently maps /...ink/ to [...ynk].

Attested cases of opacity are not like this. The opaque process in Ojibwa is not different in kind from the transparent processes in other languages. The opaque processes that we actually find in languages are natural. In OT terms, this means that those processes should devolve from standard markedness-over-faithfulness rankings, since such rankings, combined with the universality of CON, *are the only means within this theory for explaining phonological* naturalness. This result establishes a strong precondition for the adequacy of any theory of opacity in OT.

2.4 What have we learned?

Perhaps the most striking result of this review of previous work on opacity is the central role played by structure that is not present in either underlying or surface representations. With the exception of contrast preservation (§2.3.4.4), all reasonably complete theories of opacity make crucial reference, via rules or constraints, to some nonunderlying, nonsurface representation. In *SPE* (§2.2.3), it is the intermediate step of a serial derivation with ordered rules. In Stratal OT (§2.3.4.2), it is the output of one stratum that is also the input to the next stratum. In sympathy, targeted constraints, and comparative markedness (§2.3.4.3), it is a more faithful candidate than the actual output form. And in Containment or Turbidity (§2.3.4.1), the counterpart to the intermediate derivational step is not exactly a level of representation, but it is nonetheless present as coexistent, unpronounced structure in the output form.

The discussion of local conjunction (§2.3.4.1) and contrast preservation (§2.3.4.4) emphasizes another point: opacity is a result of process interaction. Our understanding of opacity cannot be separated from our understanding of what 'process' means in OT, nor can it be separated from our understanding of how different processes may make inconsistent demands on phonological mappings. Opacity is deeply connected with the phonology of a language, and any adequate theory of opacity must recognize this.

From all this, I conclude that there is some fundamental truth to the derivational view of opacity. But OT denies the existence of rules and therefore of rule ordering, for some very good reasons (McCarthy 2002b: chapter 3, Prince and Smolensky 2004: chapters 2–5, 9). We have also seen that previous attempts to meld OT with serial derivations or their analogues have not been fully successful (§2.3.4.1–§2.3.4.3). The challenge, then, is to make use of the derivational insight without losing hold of OT's essential properties and basic results. The next chapter presents a proposal intended to do exactly that.

Notes

1 The derivation in (2-4) also includes a bleeding interaction: deletion bleeds raising in the derivation of [ʃarbat].
2 Lexical exceptions will also look like they should have undergone a process but did not. Kiparsky's definition of opacity is limited to the effects of process interaction and not exceptionality. See Laferriere (1975) for a useful distinction between 'internal' opacity (exceptionality) and 'external' opacity (2-6).
3 Hayes's (1990) notion of 'precompilation' provides a way of lexically listing morphologized alternations that occur in external sandhi. Precompilation is therefore intended for phenomena like morphological mutation. It is by intent and by design inappropriate for examples like high-vowel deletion in Bedouin Arabic, where there are no morphological or grammatical conditions.
4 Oh, the grand old Duke of York, / He had ten thousand men; / He marched them up the hill, / And he marched them down again.
5 See Pater (2000) for an analysis of these data in OT using output-output correspondence.
6 This formulation is due to Jane Grimshaw.
7 The qualification 'at least' implies that there may be more than one fully faithful candidate from a given input. This is possible if there are dimensions along which candidates may differ that are not protected by faithfulness constraints. Syllabification is the standard example. For further discussion, see McCarthy (2002a, 2003c) and §3.2.4.1.
8 Throughout, I follow Prince (2002) in using comparative tableaux. The winning candidate appears to the right of the arrow, and losers are in the rows below it. Subscripted integers stand for the number of violation marks incurred by a candidate, replacing the familiar strings of asterisks. In loser rows, the effects of the constraints are indicated by W and L, W if the constraint favors the winner and L if it favors the loser. These annotations are much more useful and perspicuous than the exclamation point and shading that they replace. For example, the sufficiency of any tableau can be easily checked: every L must be outranked by (≈ to the right of) some W in the same row, and every loser row must contain at least one W.
9 I am grateful to Adam Werle for help with the Nuuchahnulth data.

10 There is a great deal of previous literature discussing the challenges that opacity presents to classic OT, including Archangeli and Suzuki (1996, 1997), Black (1993), Booij (1997), Cho (1995), Clements (1997), Chomsky (1995), Goldsmith (1996), Halle and Idsardi (1997), Idsardi (1998), Jensen (1995), Kager (1997, 1999b), McCarthy (1996, 1999), McCarthy and Prince (1993b), Noyer (1997), Paradis (1997), Prince and Smolensky (2004), Roca (1997b), Rubach (1997), and various contributions to Hermans and van Oostendorp (eds) (1999).
11 The double vertical line in the middle of tableau (2-16) is used to separate blocks of constraints that cannot be ranked on the basis of the information provided. That is, no ranking relations are asserted across this double line.
12 On compensatory lengthening, see among others Hayes (1989) and Wetzels and Sezer (eds) (1986).
13 The constraint *Complex-Coda (*Comp-Coda) is violated by any tautosyllabic syllable-final cluster.
14 The problems of constraint conjoinability and domains arise regardless of whether we regard conjoined constraints as literally present in universal Con or merely immanent in it (see Ito and Mester 2003a: 24 on this distinction). Either way, linguistic theory is obliged to explain why certain logical possibilities do not occur.
15 Modular, serial implementations of OT along the lines of the theory of Lexical Phonology have been proposed or discussed in the following works, among others: Bermúdez-Otero (1999, forthcoming), Cohn and McCarthy (1994/1998), Hale and Kissock (1998), Hale, Kissock, and Reiss (1998), Ito and Mester (2001, 2003b, 2003c), Kenstowicz (1995), Kiparsky (2003, to appear), McCarthy (2000b), McCarthy and Prince (1993b), Orgun (1996b), Potter (1994), Rubach (2000), and many of the contributions to Hermans and van Oostendorp (eds) (1999) and Roca (ed.) (1997a).
16 The hypothesis that strata can eliminate the need for extrinsic ordering is pursued in a somewhat different theoretical context by Goldsmith (1993a) and Lakoff (1993). Their theories are distinct from both LP, because they do not allow within-stratum ordering, and from Stratal OT, because they employ 'two-level' rules that can refer to input environments. (Stratal OT's markedness constraints, like classic OT markedness constraints, can only refer to output environments.)
17 *DNA* stands for 'does not apply'.
18 The processes in (2-28), though opaque, are nearly exceptionless. I know of no lexical exceptions to stressing of final closed syllables or deletion of final [ʔ]. The only lexical exceptions to epenthesis in final clusters are the words [neːrd] 'nard' (Canticles 4, verse 14), a borrowing from Persian, and [qoʃtˤ] 'truth' (Proverbs 22, verse 21), a *hapax legomenon*.
19 There is a fairly extensive literature on sympathy theory and its applications (Bakovic 2000, Davis 1997a, 1997b, de Lacy 1998, Dinnsen *et al.* 1998, Fukazawa 1999, Harrikari 1999, Ito and Mester 1997a, 1997b, 1998, Jun 1999, Karvonen and Sherman [Ussishkin] 1997, 1998, Katayama 1998, Kikuchi 1999, Lee 1999, McGarrity 1999, Merchant 1997, Odden 1997, Parker 1998, Sanders 1997, Walker 1998, 2003, Wilbur 1998).