J. Linguistics **37** (2001), 503–525. © 2001 Cambridge University Press
DOI: 10.1017/S0022226701001062 Printed in the United Kingdom

The puzzle-puddle-pickle problem and the Duke-of-York gambit in acquisition[1]

DANIEL A. DINNSEN,
KATHLEEN M. O'CONNOR & JUDITH A. GIERUT

Indiana University

(Received 4 July 2000; revised 30 March 2001)

Two classic and previously unrelated problems are reconsidered for their implications for optimality theory and acquisition. The puzzle-puddle-pickle problem centers on the debate over children's underlying representations and the characterization of interacting error patterns which, when lost, result in overgeneralizations. In response to the challenges that this problem poses, an optimality theoretic solution is offered that appeals to the second problem, the Duke-of-York gambit, which involves co-occurring generalizations with reverse effects. The solution avoids language-specific restrictions on input representations and characterizes the loss and introduction of errors by one mechanism. New insight is offered for when overgeneralization is (not) expected to occur.

1. Introduction

A long-standing problem in phonological acquisition has been the determination of the substance of children's internalized underlying representations. The standard assumption has been that children's underlying representations are target-appropriate, even in the early stages of acquisition when there may be many production errors. Some of the most compelling evidence in support of this assumption has come from Smith's (1973) diary study of his son, Amahl. However, Macken's (1980) reanalysis of certain of Amahl's interacting error patterns and his subsequent overgeneralization errors forced a reconsideration of this assumption. The circumstances of this case have come to be known as the puzzle-puddle-pickle problem due in large part to the 'chain shift' that resulted from the interaction of the error patterns at an early stage of development. One error pattern (Stopping) replaced fricatives with stops (for example, *puzzle* type words were realized

[1] We are especially grateful to Stuart Davis, Gregory Iverson, Andrew Koontz-Garboden, Laura McGarrity and two anonymous *JL* referees for their comments on a prior draft of this paper. The paper has also benefited from comments received during presentations at the 6th Mid-Continental Workshop on Phonology at Ohio State University, October 2000, and at the Annual Meeting of the Linguistic Society of America, January 2001. This work was supported in part by a grant from the National Institutes of Health DC01694 to Indiana University.

as *puddle* type words). The other error pattern (Velarization) replaced coronals with velars before a liquid consonant (for example, *puddle* type words were realized as *pickle* type words). What makes this an example of a chain shift is that the stops corresponding to target fricatives resisted the Velarization error pattern. The other part of this problem relates to a later stage of development when these error patterns were lost but new overgeneralization errors were introduced. Target *pickle* words, which had previously been produced correctly, changed to incorrect productions at the later stage, being realized as *puddle* words. The combined elements of this problem have led to an acknowledgment that at least some underlying representations for some children may be internalized incorrectly (relative to the target system).

This issue is reflected in many other case studies and has received considerable attention within rule-based derivational theories (for a review, see Dinnsen 1999). Surprisingly, however, work conducted within the constraint-based approach of optimality theory (Prince & Smolensky 1993, McCarthy & Prince 1995) has yet to consider this problem and has largely accepted without argument the standard assumption about children's underlying representations (e.g. Demuth 1995, Smolensky 1996b, Hale & Reiss 1998, Hayes 1999). The puzzle-puddle-pickle problem does, however, bear on several basic tenets of optimality theory and warrants reconsideration within this newer framework. While derivational theories can provide for children's incorrectly internalized underlying representations by imposing language-specific restrictions on those representations, no such restrictions are available within optimality theory, especially given the principle of RICHNESS OF THE BASE (e.g. Prince & Smolensky 1993, Smolensky 1996b). This principle maintains that the set of input representations is universal and thus the same for all languages and, by extension, for all children. In optimality theory, the same effects must be achieved by the constraint hierarchy. Additionally, the error patterns associated with the puzzle-puddle-pickle problem interacted in a way that yielded a chain shift, obscuring (or rendering opaque) one of the central generalizations of Amahl's phonology. A generalization can be considered opaque if, for example, it is not surface true. The characterization of opacity effects has challenged optimality theory and remains controversial (cf. Itô & Mester 1999, McCarthy 1999b, Goldrick 2000, Kiparsky 2000). Finally, the observed overgeneralization errors of this case would also seem to be at odds with optimality theoretic predictions about the course of acquisition and assumptions about constraint demotion (e.g. Tesar & Smolensky 1998). The problem posed by overgeneralization errors is why words that are produced correctly at one point in time should change to incorrect productions at a later point in time. It is not immediately obvious what evidence would motivate a child to make such a change. Given the theoretical machinery that is (and is not) available within optimality theory, the problem is to explain how such acquisition facts would be accounted for. The purpose of this paper is to reconsider these issues in the context of the general puzzle-puddle-pickle problem and to determine whether an optimality theoretic account can be formulated that is consistent with the facts and the general assumptions of acquisition. It will be argued that optimality theory offers a unified account while also offering new insight into when overgeneralization is and is not expected to occur in acquisition. The solution takes advantage of an insight from the Duke-of-York gambit (Pullum 1976, cf. McCarthy 1999c). In derivational terms, the Duke-of-York gambit is characterized by two ordered rules whereby the second rule reverses the effect of the first rule. In optimality theoretic terms, the Duke-of-York gambit would obtain from the ranking of two contextually overlapping but conflicting markedness constraints. While viable Duke-of-York effects have been documented for fully developed languages, there has always been some question about how or why such effects might arise. By connecting the Duke-of-York gambit with acquisition phenomena, we begin to arrive at an answer.

The paper is organized as follows. In section 2, a summary sketch of the puzzle-puddle-pickle problem is presented in rule-based derivational terms. Section 3 develops an optimality theoretic account of the facts consistent with general assumptions about acquisition. The solution is shown to require no language-specific restrictions on the set of available input representations and to provide for the loss of error patterns and the introduction of new overgeneralization errors by one and the same mechanism, namely the minimal demotion of markedness constraints, with the reranking motivated on the basis of positive evidence alone. In an effort to further evaluate our account, section 4 considers a similar, but contrasting, set of acquisition phenomena which appears to be resistant to overgeneralization. Some crucial factors are identified which distinguish those cases that are susceptible to overgeneralization from those that are not. As a result of these considerations, the Duke-of-York gambit is found to play a significant role in acquisition. The paper concludes with a brief summary.

2. The puzzle-puddle-pickle problem

The puzzle-puddle-pickle problem has been associated most notably with Amahl's phonological development as described by Smith (1973) and Macken (1980). The problem does, however, represent a more general constellation of pervasive issues in both first- and second-language acquisition. Those issues include the nature of underlying representations, the emergence and loss of opacity effects, and the introduction of overgeneralization errors. For a review of the evidence motivating children's incorrectly internalized underlying representations, see Dinnsen (1999). A range of opacity effects has also been documented for children with normal or delayed phonological development (e.g. Dinnsen, McGarrity, O'Connor

& Swanson 2000, Dinnsen & O'Connor 2001). In addition, the phenomenon of overgeneralization is equally common in the developing phonologies of first-language learners (normal or disordered) (e.g. Smith 1973, Menn 1983, Stoel-Gammon & Stemberger 1994, Dinnsen & McGarrity 1999, Powell, Miccio, Elbert, Brasseur & Strike-Roussos 1999). Overgeneralization errors are sometimes referred to as recidivisms, regressive overgeneralizations or hypercorrections. Some of these same issues have been documented in second-language acquisition (e.g. Major 1987, Eckman & Iverson 2000). Thus, while the case study of Amahl is highlighted here, we intend the discussion to apply to phonological development generally. For this reason, and to allow a sharper focus on the theoretical issues, we will concentrate on the more central aspects of Amahl's case.

According to Smith (1973), Amahl exhibited two error patterns that interacted to yield opaque outputs during his early stages of development (age 2;2-2;11). One of the error patterns, which we dub Stopping, replaced fricatives with stops in a context-free fashion. As a result, target words such as *puzzle* were realized as *puddle*, as illustrated in (1a). The other interacting error pattern, which we identify as Velarization, replaced target coronal stops with a velar before /l/, possibly as a result of assimilation. As shown in (1b), target *puddle* words were realized as *pickle* words.[2] Coronal stops occurred elsewhere and thus were unaffected by this process in those other contexts. Additionally, target *pickle* words were realized with an appropriate velar stop, as can be seen in (1c).

(1) Amahl (age 2;2-2;11)
   (a) *Puzzle* words realized as *puddle* words (Stopping)
      [pʌdl̩] 'puzzle'
      [pɛntl̩] 'pencil'
      [wɪtl̩] 'whistle'
   (b) *Puddle* words realized as *pickle* words (Velarization)
      [pʌgl̩] 'puddle'
      [bɔkl̩] 'bottle'
      [hæŋgl̩] 'handle'
   (c) *Pickle* words realized target appropriately
      [pɪkl̩] 'pickle'
      [təːkl̩] 'circle'

The rules (or processes) responsible for these errors clearly interacted. If Stopping had been ordered before Velarization in a feeding relation, as

[2] Macken (1980) noted that there were some apparent exceptions to Velarization. It appears, however, that there may have been only one true exception, namely the word *little*. The relative high frequency of this word in English may have contributed to its exceptional behavior. See Gierut, Morrisette & Champion (1999) for one possible optimality theoretic account incorporating lexical frequency and neighborhood density to handle apparent exceptions of this sort.

exemplified in (2a), *puzzle* words would have been realized contrary to fact as *pickle* words. However, the product of Stopping did not undergo Velarization, motivating a counterfeeding relation with Velarization ordered before Stopping. The derivation in (2b) illustrates the attested interaction and resultant opacity. The Velarization error pattern is rendered opaque from this order of the rules because the phonetic realization of target *puzzle* words appears to violate Velarization.

(2) Sample derivations

(a) Unattested feeding order

UR	/pʌzl/ 'puzzle'	/pʌdl/ 'puddle'
Stopping	pʌdl	—
Velarization	pʌgl	pʌgl
PR	*[pʌgl̩]	[pʌgl̩]

(b) Attested counterfeeding order

UR	/pʌzl/ 'puzzle'	/pʌdl/ 'puddle'
Velarization	—	pʌgl
Stopping	pʌdl	—
PR	[pʌdl̩]	[pʌgl̩]

Various pieces of evidence led Smith (1973) to postulate target-appropriate underlying representations for these words. The evidence included the systematic correspondence between the target system and Amahl's speech, the differential behavior of target stops and fricatives before /l/, the correct production of *pickle* words, and Amahl's accurate comprehension of these words. Another piece of evidence came from a subsequent stage of development where these two rules were no longer operative (age 3;9). Rule loss is relevant because the absence or loss of a rule should allow previously affected words to be realized phonetically as they are presumed to be represented underlyingly. The Stopping rule was lost prior to the loss of Velarization and did indeed result in target-appropriate realizations of *puzzle* words, as shown in (3a). The loss of the Velarization rule resulted in correct realizations of *puddle* words, as shown in (3b). Such facts would seem to support the claim that Amahl had indeed internalized target-appropriate underlying representations. However, Macken (1980) in a reanalysis of the data showed that at that same point in time, Amahl exhibited new overgeneralization errors such that the previously correct *pickle* words changed, being produced incorrectly as *puddle* words, as shown in (3c).[3]

[3] This overgeneralization was not across-the-board. That is, some target velar + liquid sequences did begin to appear around that same point in time. For a possible optimality theoretic account of similar lexical diffusion effects, see Gierut et al. (1999).

(3) Amahl (age 3;9)
  (a) *Puzzle* words realized correctly (Stopping lost)
    [pʌzl̩] 'puzzle'
    [pɛnsl̩] 'pencil'
    [wɪsl̩] 'whistle'
  (b) *Puddle* words realized correctly (Velarization lost)
    [pʌdl̩] 'puddle'
    [bɔtl̩] 'bottle'
    [hændl̩] 'handle'
  (c) *Pickle* words realized incorrectly as *puddle* words (Overgeneralization)
    [pɪtl̩] 'pickle'
    [səːtl̩] 'circle'
    [wɪntl̩] 'winkle'

Derivational theories have available a number of different mechanisms which could in principle account for the introduction of overgeneralization errors. One possibility might have been that a (dissimilatory) Fronting rule was added to the grammar at the point the Velarization rule was lost. Such a rule might have converted velars into coronals when followed by a liquid. It is admittedly unclear what facts would have motivated Amahl to add such a rule to his grammar, especially while at the same time losing Velarization. Another possibility is that both Velarization and Fronting might have co-existed in Amahl's grammar during both stages with the two rules reordering over time. The co-occurrence of the rules would have the consequence that coronals and velars would have acted the same before liquid consonants, although with different results depending on the order of rules. The early stage might have had Fronting ordered before Velarization, and the later stage might have had Velarization ordered before Fronting. Either order of the rules would have constituted a Duke-of-York derivation with the later rule reversing the effect of the earlier rule (Pullum 1976). While such an account might be possible within derivational theories, it is acknowledged that there would be no empirical support for the first of the two ordered rules in either stage of development. Moreover, the change in rule ordering would not follow from any known principle of rule reordering (e.g. King 1973). We will see later that a Duke-of-York solution takes on added significance within optimality theory. The option more closely aligned with Macken's account is that the overgeneralization errors were taken as evidence that the *pickle* words were internalized incorrectly as *puddle* words from the outset (possibly as a result of misperception). That is, during the early stages, the active Velarization rule would have hidden the erroneous underlying representations of *pickle* words, which would have become evident only when the Velarization rule was lost. The consequence is that although *pickle* words might have been produced correctly during the earlier stages, those correct productions would have come about from incorrectly internalized representations and the operation of the Velarization rule. This would have constituted a case of 'correct productions for the wrong reasons' (Dinnsen 1999).

Incorrectly internalized representations of this sort are not difficult to model within derivational theories. In fact, one of the ways in which grammars are expected to differ is in the substance of their underlying representations. It thus might have been reasonably argued that (some of) Amahl's underlying representations during the early stages differed from those of the target system by being simpler in the number and type of available contrasts. For example, velar + liquid sequences might have been banned from Amahl's underlying representations, possibly as a result of a morpheme-structure condition or underspecification. The Velarization rule would have converted (or filled in) those simplified, undifferentiated (incorrect) underlying representations into their corresponding phonetic representations with a velar + liquid. While the subsequent loss of the Velarization rule would seem to have been a step in the right direction for Amahl, leading to the correct production of *puddle* words, it was at that point that the new overgeneralization errors were introduced. Those new errors were presumably not the result of a rule, but rather the result of the earlier simplified and incorrect underlying representations. Eventually, Amahl would come to produce *pickle* words correctly, as his underlying forms for those words were restructured in conformity with the target system. That restructuring would presumably have been guided by his more accurate recognition of the properties of specific words.

Such an account serves to instantiate a prediction about acquisition within derivational theory, namely that overgeneralization of this sort is expected when underlying forms are incorrectly internalized and a rule is lost. The following section considers how these same facts can be accounted for within optimality theory, which has no rules (and thus no rule loss) and which disallows language-specific restrictions on the set of underlying representations.

3. An optimality theoretic account

Optimality theory differs from derivational theories in several important respects. There are no rules and thus no rule ordering relationships, no derivations, no intermediate levels of representation, and no language-specific restrictions on the set of underlying representations. Instead, for any given input (underlying representation), a ranked set of universal constraints evaluates in parallel a potentially infinite set of candidates and selects one as optimal. The optimal candidate is the one that best satisfies the constraint hierarchy. Constraints are of at least two fundamental and often antagonistic types, namely markedness constraints and faithfulness constraints.

Markedness constraints are formulated exclusively in terms of output properties and militate against marked segment types, sequences and structures. Faithfulness constraints, on the other hand, demand identity between input and output strings. The many production errors that occur in early stages of development have been characterized by the default ranking of markedness constraints over the antagonistic faithfulness constraints (Gnanadesikan 1996, Smolensky 1996a). The process of acquisition leading to target-appropriate realizations is presumed to proceed by the reranking of constraints, specifically by the minimal demotion of markedness constraints (Tesar & Smolensky 1998).

In formulating an optimality theoretic account of the facts of Amahl's system, we begin with the error patterns and their interaction during the early stages. The Stopping error pattern affected the manner feature [continuant] and is suggestive of an antagonism between the markedness and faithfulness constraints given in (4).

(4) Constraints and ranking for the Stopping error pattern
*FRIC: Avoid fricatives
ID[manner]: Corresponding segments must be identical in terms of manner features
Ranking: *FRIC ≫ ID[manner]

By ranking the markedness constraint *FRIC over the faithfulness constraint ID[manner] as shown above, the claim is made that it is more important to avoid fricatives than it is to be faithful to a segment's input manner feature. Fricatives would thus be banned from phonetic outputs, no matter what might be assumed about the input representation.

The Velarization error pattern involved a change in [place] features and is suggestive of two other antagonistic constraints, given in (5).

(5) Constraints and ranking for the Velarization error pattern
*dl: Avoid coronals before liquid consonants
ID[place]: Corresponding segments must be identical in terms of place features
Ranking: *dl ≫ ID[place]

The motivation for *dl may find some basis in the English prohibition of tautosyllabic coronal stop + liquid consonant sequences (e.g. Davis 1991). It may also be related to conceptions of feature geometry which attempt to account for the tendency of consonants to take on (and even promote) the secondary place features of adjacent consonants (e.g. Clements & Hume 1995). The idea here would be that the secondary dorsal feature of the velarized /l/ would be associated with the preceding coronal consonant and promoted to a primary place feature. In any event, a coronal consonant before a velarized /l/ would violate the constraint, and the ranking in (5) would compel the coronal consonant to change to a velar in that context.[4] An input velar consonant before /l/ would not violate *dl and would thus be realized faithfully.

While these constraints and rankings correctly predict that target *puddle* words would be realized as *pickle* words, they also make the empirically incorrect prediction that target *puzzle* words would be realized as *pickle* words. This latter point is illustrated by the tableau in (6). For simplicity's sake, we will ignore irrelevant phonetic differences among competing output candidates and will thus limit our choices to the words *puzzle*, *puddle* and *pickle* throughout the tableaux as the most relevant and representative candidates. Candidate (b), which should win on empirical grounds, is eliminated by its fatal violation of *dl.

(6) *Puzzle* words realized as *pickle* words (unattested)

puzzle	*Fric	*dl	ID[manner]	ID[place]
(a) puzzle	*!	*		
(b) puddle		*!	*	
☞(c) pickle			*	*

Legend for tableaux:

☞ = optional output   ☛ = erroneous prediction   * = constraint violation

*! = fatal violation   ⋮ or, = equal ranking | or >> = crucial ranking

The problem is that no single ranking of these constraints can account for the combined set of facts in (1). The need to realize *puzzle* words as *puddle* words while target *puddle* words are realized as *pickle* words is at the heart of the problem. The observed substitutions participate in a chain shift and result in a type of opacity, namely a non-surface-true generalization. The characterization of such opacity effects, especially those associated with chain shifts, has been argued to require the local conjunction of faithfulness constraints (e.g. Smolensky 1993, Kirchner 1996).[5] Along these lines, it can

[4] Even if *dl were found to have no basis in assimilation, an independently necessary and highly ranked faithfulness constraint demanding identity between an input and an output in terms of the feature [labial] would prevent the coronal from being replaced by a labial consonant. The evidence for this ranking is that labial consonants were neither replaced by other places of articulation nor were they the substitute for other consonants.

[5] For other proposals about the characterization of opacity effects and their emergence in acquisition, see Itô & Mester (1999), McCarthy (1999b), Dinnsen, McGarrity, O'Connor & Swanson (2000), Goldrick (2000) and Dinnsen & O'Connor (2001).

be observed that *pickle* as an output correspondent of input *puzzle* only violates the two low-ranked faithfulness constraints ID[place] and ID[manner]. The dominated character of these constraints renders them ineffectual in ruling out *pickle* as the output for input *puzzle*. However, by locally conjoining those two constraints into one single complex constraint and ranking the complex constraint above the two elementary constraints, the desired result can be achieved. That is, *pickle* words can be eliminated as correspondents of *puzzle* words, but be preferred as correspondents of *puddle* words. A locally conjoined constraint is considered violated if and only if both of the component constraints are violated in the same domain (in this instance, in the same segment). The ranking of the complex constraint over the elementary constraints is presumed to be universal, following from the ELSEWHERE CONDITION and the special/general relation that holds among these constraints (e.g. Kiparsky 1973). The constraints and ranking in tableau (7) illustrate part of the effect for input *puzzle*. The faithful candidate (a) complies with the locally conjoined constraint but incurs one violation for each of the other equally ranked constraints *FRIC and *dl, the second of which is fatal. Candidates (b) and (c) each incur one violation of an undominated constraint, with the choice being passed down to the lower ranked faithfulness constraints. Candidate (c) is eliminated because it violates both of the faithfulness constraints, while candidate (b) only violates ID[manner].

(7) Locally conjoined constraint and tableau for *puzzle* realized as *puddle*
ID[manner]&ID[place] (LC): Corresponding segments must be identical in terms of either [place] or [manner] features

puzzle	LC	*FRIC	*dl	ID[manner]	ID[place]
(a) puzzle		*	*!		
☞ (b) puddle			*	*	
(c) pickle	*			*	*!

The tableau in (8) shows how input *puddle* words would be realized as *pickle* words with these constraints and rankings. Candidate (a) violates both *FRIC and *dl, either of which would be fatal. The faithful candidate (b) complies with LC and *FRIC but fatally violates *dl. Candidate (c) does not violate the locally conjoined constraint because, once again, only one part of the constraint (in this instance, ID[place]) is violated. Candidate (c) thus survives as optimal, only violating one low-ranked faithfulness constraint.

(8) *Puddle* realized as *pickle*

puddle	LC	*Fric	*dl	ID[manner]	ID[place]
(a) puzzle		*!	*	*	
(b) puddle			*!		
☞ (c) pickle					*

The tableau in (9) for *pickle* words completes the account of Amahl's early stages.

(9) *Pickle* realized as *pickle*

pickle	LC	*Fric	*dl	ID[manner]	ID[place]
(a) puzzle	*!	*	*	*	*
(b) puddle			*!		*
☞ (c) pickle					

We now turn to the account of the later stage, where the two error patterns were no longer evident but overgeneralization occurred. The loss of the Stopping error pattern would be attributed to the demotion of *Fric below ID[manner]. That ranking would ensure that input *puzzle* words would be realized faithfully. Similarly, the demotion of *dl below ID[place] would result in the loss of the Velarization error pattern, allowing input *puddle* words to be realized faithfully. In accord with our earlier assumption, the locally conjoined constraint would continue to outrank the elementary faithfulness constraints that comprise it. If, as in our derivational account above, *pickle* words had been internalized incorrectly as *puddle* words, they too would have been realized as *puddle* words, consistent with the overgeneralization facts. The tableaux in (10) for *puzzle* words and in (11) for both *puddle* and *pickle* words illustrate these points.

(10) *Puzzle* realized as *puzzle*

puzzle	LC	ID[manner]	ID[place]	*Fric	*dl
☞ (a) puzzle				*	*
(b) puddle		*!			*
(c) pickle	*!	*	*		

(11) *Pickle* and *puddle* realized as *puddle*

puzzle	LC	ID[manner]	ID[place]	*Fric	*dl
(a) puzzle		*!		*	*
☞ (b) puddle					*
(c) pickle			*!		

The problem with this account of the *pickle* words and their associated overgeneralization is that it is at odds with richness of the base. That is, if we were to disallow velar + liquid sequences as possible underlying representations, we would be imposing a language-specific restriction on the set of available input representations. Richness of the base provides that the set of input representations is universal, precluding any such language-specific restrictions. To be consistent with this principle, we would need to admit velar + liquid sequences as possible underlying sequences. The phonetic non-occurrence of these sequences must follow from the constraint hierarchy. Consequently, if we were to now allow the *pickle* words to be represented target appropriately, then some previously inactive, but universal markedness constraint must become active and dominate ID[place] to induce the change from a *pickle* word to a *puddle* word. Let us suppose that the relevant markedness constraint is that given in (12).

(12) *gl: Avoid velar consonants before liquid consonants

As defined, this contextual markedness constraint directly conflicts with *dl. The motivation for the constraint may find some basis in dissimilation or the need to avoid OCP violations or a dispreference for multiply linked features. The substantive details of the constraint may ultimately prove to differ somewhat from our formulation. However, the important point is that its context overlaps with that of *dl. The overgeneralization facts and the conflict between *dl and *gl reveal another example of opacity in this child's system. The effect of this conflicting and presumably universal constraint would not have been evident in the earlier stage due to its lower ranking in the hierarchy, especially if it had been ranked below ID[place]. Given that velars could occur in other contexts during both stages, the problem was apparently a matter of positional markedness. A tentative constraint ranking consistent with the facts of this later stage is given in (13).

(13) Tentative constraint rankings for the two stages[6]
  (a) Early stage: LC, *Fric, *dl ≫ ID[manner], ID[place] ≫ *gl
  (b) Overgeneralization stage: LC, *gl ≫ ID[manner], ID[place] ≫ *Fric, *dl

[6] Other rankings of these constraints are also consistent with the facts. The ranking arguments will be refined below, but even then, certain rankings will remain indeterminate.

One of the important differences in the account of the overgeneralization stage (relative to the prior stage) is the ranking of *gl over ID[place], which in turn dominates *dl. While we would be able to account for the facts of the two stages in isolation with these different constraint rankings, we would be doing so in a way that is at odds with other optimality theoretic predictions about the course of acquisition and assumptions about constraint demotion (Tesar & Smolensky 1998). That is, the accounts of the two stages do not provide for the transition from one stage to the next. If children's initial state is to be characterized by the dominance of markedness constraints over faithfulness constraints (e.g. Hayes 1999, McCarthy 1999a), acquisition should proceed by the minimal demotion of markedness constraints until conformity with the target system is achieved. Once a faithfulness constraint has come to dominate a markedness constraint (as tentatively suggested by our account of the earlier stage with ID[place] dominating *gl), the questionable use of negative evidence would be required to motivate the subsequent demotion of faithfulness below markedness (as entailed in our tentative account of the later stage). An example of negative evidence would be a child's reliance on the observation that some sound or sound sequence did not occur. In this case, Amahl would have had to observe that velar + liquid sequences did not occur in the primary linguistic data to which he was exposed. Such an observation would be unlikely or at least surprising given the facts of English and of his earlier stage of development. It is generally assumed that children require positive evidence for learning (or in this case for constraint reranking). An example of positive evidence would be the recognition that some sound or sound sequence can occur. We thus would not expect markedness constraints to be promoted over faithfulness constraints given that no positive evidence would be available to motivate such a reranking.

To see the specific problem that our tentative account poses, consider again Amahl's earlier stage of development. At that point, *dl was active (dominating ID[place]), and *gl was inactive (presumably due to its being dominated by ID[place]). Such a ranking would be consistent with incorrect productions of *puddle* words and correct productions of *pickle* words. Compare that with the ranking in our account of the later stage of development where *puddle* words were produced correctly and *pickle* words were produced incorrectly. The inactive character of *dl during that later stage would be consistent with its being dominated by ID[place]. The essential difference between the two stages on this point is given in (14).

(14) Tentative rankings for the two stages
  (a) Early stage: *dl $\gg$ ID[place] $\gg$ *gl
  (b) Later stage: *gl $\gg$ ID[place] $\gg$ *dl

The reranking of these constraints entails two changes of note. One of the changes, namely the demotion of *dl below ID[place], is in part consistent with general assumptions about constraint rerankings and might be expected if coronal+liquid sequences were to be produced correctly, as did in fact occur for Amahl in the later stage. The other change, however, the promotion of *gl over the previously dominant faithfulness constraint, is entirely unexpected. No positive evidence would ever be available to motivate the demotion of a previously dominant faithfulness constraint below a markedness constraint.

The solution to this problem lies in the recognition that some target-appropriate productions can arise and be judged optimal even though the ranking does not conform to the target ranking. For this to be so, a particular faithfulness constraint must not dominate any of the relevant markedness constraints. That is, some correct productions would be predicted to occur during both stages of development if neither of the two markedness constraints, *dl and *gl, were dominated by ID[place] during those two stages. The difference between the two stages would be reflected in the ranking of the two markedness constraints relative to one another, as shown in (15).

(15) Relative ranking of markedness constraints
    (a) Early stage:[7] *dl ≫ *gl ≫ ID[place]
    (b) Later stage: *gl ≫ *dl, ID[place]

By ranking and reranking these two markedness constraints in this way, the claim is made that coronals and velars are both disfavored before liquids, but one of those sequences is worse than the other, depending on the stage of development and the relative ranking of the two markedness constraints. Faithfulness plays little or no role in choosing between competitor candidates in this case. The dominance of one markedness constraint over the other eliminates the only other viable candidate, yielding a winner that violates a markedness constraint that is itself not dominated by a faithfulness constraint. In some sense, the winner is the best of the offending candidates. The abbreviated tableaux in (16) illustrate this point for the early stage.

(16) Early stage
    (a) *Puddle* realized as *pickle*

puddle	*dl	*gl	ID[place]
(a) puddle	*!		
☞ (b) pickle		*	*

[7] The ranking of *dl over *gl in the early stage may follow from a default preference for assimilation over dissimilation.

(b) *Pickle* realized as *pickle*

pickle	*dl	*gl	ID[place]
(a) puddle	*!		*
☞ (b) pickle		*	

The most minimal demotion of *dl below *gl in the later stage would have the consequence of installing *dl in the stratum with ID[place] and would have been motivated on the basis of positive evidence, namely Amahl's recognition that coronal stop + liquid sequences could occur (independent of corresponding inputs).[8] The tableaux in (17) show how the overgeneralization effect is achieved in the later stage of development with this new reverse ranking of the markedness constraints.

(17) Overgeneralization effect in later stage of development
(a) *Puddle* realized as *puddle*

puddle	*gl	*dl	ID[place]
☞ (a) puddle		*	
(b) pickle	*!		*

(b) *Pickle* realized as *puddle*

puddle	*gl	*dl	ID[place]
(a) puddle		*	*
☞ (b) pickle	*!		

A consequence of this account is that no further overgeneralization is possible beyond this overgeneralization stage. That is, coronal + liquid sequences are prevented from subsequently overgeneralizing back to velar + liquid sequences given the rerankings that are available. More specifically, on the basis of positive evidence alone (that is, observing the occurrence of velar + liquid sequences), *gl would have to be demoted to the stratum below ID[place] (and *dl). The resultant ranking would yield target-appropriate realizations of coronals and velars before liquids without the possibility of overgeneralization.

It is also important to note that no other sequence of events or constraint rerankings would have been motivated given the early stage rankings in

[8] Recall that the Stopping error pattern ceased prior to the loss of Velarization. *Puzzle* words were no longer being produced as *puddle* words.

(15a). That is, to demote *gl below ID[place] while retaining the undominated ranking of *dl would be empirically indistinguishable from the early stage ranking and result in no change in pronunciation. The only reranking that could have empirical consequences is the one that occurred, namely that shown in (15b). The demotion of *dl rather than *gl is, in fact, predicted by the constraint demotion algorithm for acquisition (Tesar & Smolensky 1998). Demotion of a constraint is motivated by mismatches between the outputs of the child's system and those of the adult system. In this case, demotion of *dl would have occurred when Amahl observed that his outputs for input *puddle* words, which were produced as *pickle* words at the early stage, differed from the adult outputs for the same words. According to the algorithm, when a mismatch is observed, the child demotes the highest ranked constraint that the target-appropriate candidate violates, in this case *dl, to the stratum immediately below the highest ranked constraint that the child's output violates, that is, *gl. This would result in the ranking for the overgeneralization stage. Turning to target *pickle* words at the early stage, there was no mismatch between Amahl's outputs and those of the target system; all were produced correctly as *pickle* words. In the absence of a mismatch, constraint demotion is not motivated because there is no evidence that the child's current ranking is incorrect.

The solution that we have settled on accounts for the facts of the various stages (including the opacity effects of the early stage, the overgeneralization effects of the subsequent stage, and the absence of overgeneralization in the later stages) along with the transition from one stage to the other in a way that is consistent with assumptions about richness of the base and constraint demotion. The constraint rankings for the various stages are summarized in (18).

(18) Constraint rankings for the stages of development
- (a) Early stage: LC, *Fric, *dl ≫ *gl ≫ ID[manner], ID[place]
- (b) Overgeneralization stage: LC ≫ ID[manner], *gl ≫ *dl, *Fric, ID[place]
- (c) Later stages (adult English): LC ≫ ID[manner] ≫ *dl, *Fric, ID[place] ≫ *gl

Returning to the issue of Amahl's underlying representations, one of the ways in which our optimality theoretic account differed from the derivational account was that it did not need to appeal to incorrectly internalized representations. On the contrary, it provided for the possibility of target-appropriate underlying representations throughout Amahl's stages of development. However, the target-appropriate realizations of *pickle* words during the early stage and of *puddle* words during the overgeneralization stage had little to do with the substance of Amahl's underlying representations; rather, those target-appropriate productions followed from the constraint hierarchy, specifically the ranking of the two opposing markedness constraints, *dl and *gl, relative to one another with neither being dominated by the faithfulness constraint ID[place]. The explanation was exactly the same for the target-inappropriate realizations of *puddle* words (Velarization) during the early stage and of *pickle* words during the overgeneralization stage. Again, our account depended little on the substance of Amahl's underlying representations at either stage of development. It is thus predicted that some correct and some incorrect productions will likely follow from highly ranked markedness constraints. This raises a number of interesting questions: Are there other similar cases in acquisition where dominant markedness constraints yield both correct and incorrect productions? If so, are they equally vulnerable to overgeneralization? The following section considers these questions with the intent of revealing when overgeneralization is and is not expected to occur.

4. A comparison with other acquisition phenomena

An informative comparison can be made by considering children's early error patterns involving allophonic phenomena. Smith (1973) provides relevant and representative data again from Amahl (cf. Dinnsen 1996). During his early stages of development (age 2;2), Amahl produced voiced and voiceless obstruents in complementary distribution. Simplifying somewhat, obstruents were voiced in syllable onsets and voiceless in codas.[9] This pattern resulted in some correct and some incorrect productions, depending on the context of the target sound. For example, target voiced obstruents were produced correctly in onsets, but were produced incorrectly as voiceless in codas. Conversely, target voiceless obstruents were produced incorrectly as voiced in onsets, but were produced correctly in codas. A rule-based derivational account of the complementary distribution of these sounds and the associated error patterns would invoke a rule to ensure the equivalent of devoicing in codas and voicing in onsets. Subsequently, Amahl began to produce voiced obstruents correctly in codas, with the target voice contrast first established in that context (age 2;4). Interestingly, however, no overgeneralization errors occurred then (or even later when the contrast extended to onsets (age 2;5)). In sum, target voiceless obstruents in codas never overgeneralized to become voiced, and target voiced obstruents in onsets never overgeneralized to become voiceless.

Derivational theories would have difficulty explaining the absence of overgeneralization errors in either of these contexts while also allowing overgeneralization in the puzzle-puddle-pickle case. That is, there is nothing to exclude the possibility of overgeneralization of the unattested sort because Amahl could have, for example, internalized all initial obstruents as voiceless

[9] The actual fact is that obstruents were voiceless unaspirated lenis in word-initial position, voiced unaspirated lenis in word-medial position and voiceless fortis (aspirated or unaspirated) in word-final position.

during the early stage of development. Such an assumption would have been supported by the relative unmarkedness of voiceless obstruents and the occurrence of voiceless obstruents in the primary linguistic data to which Amahl would have been exposed. To the extent that this assumption was incorrect for target voiced obstruents in that context, the rule during the early stage would have ensured the actual phonetic result that those obstruents were realized as voiced. The rule would have effectively hidden (or occulted) any assumptions (correct or incorrect) about underlying voicing. Any erroneous assumption would, however, become evident upon the loss of the allophonic rule. That is, if word-initial target voiced obstruents had been internalized incorrectly as voiceless, they would have been realized as voiceless, contrary to the actual facts. While derivational theories predict that such overgeneralization should be possible, the first-language acquisition literature fails to reveal any evidence of overgeneralization errors as phonemic contrasts are emerging from the loss of allophonic rules (see Dinnsen 1999 for a review).[10] The absence of overgeneralization errors in these cases appears to be nonaccidental and thus is in need of an explanation.

Optimality theoretic accounts of allophonic phenomena are especially relevant here because in all such cases two opposing markedness constraints must outrank an antagonistic faithfulness constraint. The issue, though, is whether the markedness constraints interact with one another. Amahl's early stage of development with allophonic voicing might have been characterized by the constraints and ranking in (19).

(19) Constraints and ranking for early stage
  (a) Markedness constraints
    *VoicedCoda: Avoid voiced obstruents in codas
    *VoicelessOnset: Avoid voiceless obstruents in onsets[11]

[10] This observation can also be extended to second-language learners as they effect a phonemic split (Eckman & Iverson 2000).

[11] This constraint (or something like it) is independent of *VoicedCoda and is required on the following typological grounds. The complementary distribution of voicing in the early stage constitutes one instance of the typology and requires both markedness constraints to be undominated. The subsequent stage with the voice contrast in codas but not onsets represents another instance of the typology and requires the demotion of *VoicedCoda below ID[voice] but the sustained dominance of *VoicelessOnset to account for the persistence of the error pattern in onsets. Languages such as German with a voice contrast in onsets and neutralization in codas represent yet another instance of the typology and require the demotion of *VoicelessOnset below ID[voice] but the dominance of *VoicedCoda. To account for that instance of the typology that is exemplified by English, both markedness constraints must be dominated by ID[voice]. The final instance of the typology would be typified by those other languages with no voiced obstruents in any context. To achieve that effect, *VoicelessOnset must be ranked below ID[voice], which in turn would be dominated by *VoicedCoda and the context-free version of that constraint (*VoicedObstruent), which would disfavor voiced obstruents elsewhere. What is excluded from this typology is the logically possible but unattested case of a language with no voiceless obstruents. This assumes that the universal constraint set does not contain a context-free constraint disfavoring voiceless obstruents. Thus, while *VoicelessOnset

(b) Faithfulness constraint
ID[voice]: Corresponding segments must be identical in terms of laryngeal features
(c) Ranking: *VoicedCoda, *VoicelessOnset ≫ ID[voice]

The two markedness constraints in this case militate against opposite values for the feature [voice], but do so in nonoverlapping complementary contexts. For this reason, no ranking relationship can be established between these two markedness constraints. Ranking either over the other would have no different empirical consequence. Nevertheless, by ranking both markedness constraints above the faithfulness constraint, the desired allophonic effect can (consistent with richness of the base) be achieved without imposing any language-specific restrictions on the set of input representations.

The subsequent stage of development, in which the voice contrast first emerged in codas, would be characterized by the conventional demotion of *VoicedCoda below ID[voice] to yield the ranking in (20). The persistence of the voicing error pattern in onsets requires the sustained dominance of *VoicelessOnset over ID[voice].

(20) Constraint ranking for later stage
*VoicelessOnset ≫ ID[voice] ≫ *VoicedCoda

The eventual extension of the voice contrast to onsets as in adult English would follow from the demotion of *VoicelessOnset below ID[voice].

The significance of this case and of our optimality theoretic account is that no ranking or reranking of the constraints could possibly result in the introduction of overgeneralization errors. Optimality theory makes the correct prediction in this instance by precluding the possibility of overgeneralization.

Our accounts of the puzzle-puddle-pickle problem and the allophonic voicing problem share a number of important characteristics, but also differ in ways that offer insight into when overgeneralization is and is not expected to occur in acquisition. Both cases had two opposing markedness constraints ranked above an antagonistic faithfulness constraint during the early stage of development. The constraints and rankings in both cases accounted for the absence of a target contrast in Amahl's speech during those early stages. That is, there was no contrast between target *puddle* words and *pickle* words, nor was there a voice contrast in any context. Also, as a result of the constraint rankings in both cases, some target words were produced correctly and others incorrectly. Another commonality was that there was no need to impose language-specific restrictions on the set of underlying representations or postulate underlying representations that were different from those of the target system. The constraint hierarchy did all of the work that might have

disfavors voiceless obstruents under certain well-defined circumstances, there is otherwise nothing marked about the context-free occurrence of voiceless obstruents.

been done by an alternative derivational theory account with language-specific restrictions on underlying representations. The account for both cases thus fully complied with richness of the base.

An important difference resided, however, in the substance of those highly ranked opposing markedness constraints. In the puzzle-puddle-pickle case, the two markedness constraints contextually overlapped such that different rankings relative to one another had different empirical consequences. For the early stage, the dominance of *dl over *gl yielded the Velarization error pattern. For the subsequent intermediate stage, the reverse ranking of these two constraints yielded the overgeneralization error pattern. These two highly ranked markedness constraints participated in a Duke-of-York gambit (McCarthy 1999c; cf. Pullum 1976). By comparison, in the allophonic voicing case, different rankings of the two markedness constraints over the faithfulness constraint could have no different empirical consequence. Because the two markedness constraints could not possibly have affected one another due to their complementary contexts, the Duke-of-York gambit was circumvented altogether. Consequently, it would follow that overgeneralization errors cannot arise from an error pattern that involves the complementary distribution of sounds. Conversely, overgeneralization errors are expected (but not required) when an error pattern is attributed to markedness constraints that contextually overlap and participate in a Duke-of-York gambit.

The different behavior (vis-à-vis the (non)occurrence of overgeneralization errors) in these two cases underscores the significance of the Duke-of-York gambit for acquisition. In his arguments for Duke-of-York derivations in fully developed languages, Pullum (1976: 100) suggested a possible connection to acquisition in his concluding paragraph as follows:

> If it is indeed true that the performance of the infant learning to speak and understand his language is something that may be appropriately modeled as a process of constructing and internalizing a fully adequate grammar of the language, we can say that although it could conceivably make the task easier, the child is not equipped with a subconscious instruction 'avoid constructing a grammar that defines Duke-of-York derivations'. We are reminded again that we have so few clear ideas about how the child is equipped that as yet we can hardly imagine how he does what he does.

To our knowledge, no acquisition evidence has ever been presented for or against the Duke-of-York gambit. The puzzle-puddle-pickle problem (along with the other documented cases of overgeneralization) would appear to be relevant to and supportive of the Duke-of-York gambit. Interestingly, the relevance of the Duke-of-York gambit for acquisition only becomes evident within optimality theory. That is, the theory's various requirements and assumptions (that is, the constraint demotion algorithm, the universality of constraints, and richness of the base) force a Duke-of-York solution. On the other hand, derivational theories have other descriptive options available (for example, rule loss, language-specific restrictions on underlying representations, rule addition and rule reordering) which would allow for a variety of rather different accounts. To the extent that optimality theory continues to require Duke-of-York solutions for those cases vulnerable to overgeneralization as distinct from those cases that resist overgeneralization, both the theory and the Duke-of-York gambit accrue added support.

5. Conclusion

The puzzle-puddle-pickle problem is seen to represent a constellation of issues relevant to the evaluation of optimality theory. One issue was the opacity that resulted from the interaction of the Velarization and Stopping error patterns during Amahl's early stages of development. This is important because the occurrence of opacity in this case contributes to the broader claim that opacity is a naturally occurring effect, even in early stages of acquisition, and must be provided for. Our account of this opacity effect appealed to the local conjunction of faithfulness constraints. Another issue was the characterization of the loss of error patterns and the introduction of new overgeneralization errors. On this point, derivational theories and optimality theory appear to make different claims. Derivational theories predict that overgeneralization errors can follow from the combined effect of incorrectly internalized underlying representations and rule loss. While some cases of overgeneralization can be accounted for in this way, many other cases of overgeneralization are predicted to be possible but seem not to occur. The problem is that derivational theories offer no explanation for the seemingly nonaccidental absence of overgeneralization errors in certain cases. Optimality theory, on the other hand, predicts that overgeneralization is possible if and only if an error pattern is governed by two opposing and overlapping universal markedness constraints which are not dominated by an antagonistic faithfulness constraint. More importantly, the two markedness constraints overlap and interact such that different rankings will have different empirical consequences (the Duke-of-York gambit). Interestingly, the conflict inherent in overgeneralization errors reveals another example of an opacity effect in acquisition. This opacity effect was characterized by the simple ranking of universal markedness constraints. Overgeneralization errors are predicted not to occur as a development of an error pattern involving sounds in complementary distribution. The basis for this latter prediction is that different rankings of the markedness constraints responsible for the complementary distribution of sounds cannot interact with one another, precluding the Duke-of-York gambit and overgeneralization. Finally, on the issue of children's underlying representations, optimality theory was found to offer a solution to the puzzle-puddle-pickle problem without imposing language-specific restrictions on the set of available input representations. Our accounts thus fully complied with the optimality theoretic principles of richness of the base and constraint demotion.

REFERENCES

Clements, G. N. & Hume, E. V. (1995). The internal organization of speech sounds. In Goldsmith, J. A. (ed.), *Handbook of phonological theory*. Cambridge, MA: Blackwell. 245–306.

Davis, S. (1991). Coronals and the phonotactics of nonadjacent constraints in English. In Paradis, C. & Prunet, J.-F. (eds.), *Phonetics and phonology* **2***: The special status of coronals: internal and external evidence*. New York: Academic Press. 49–60.

Demuth, K. (1995). Markedness and the development of phonological structure. In Beckman, J. (ed.), *Proceedings of the North East Linguistic Society* **25**. Amherst, MA: GLSA. 13–25.

Dinnsen, D. A. (1996). Context-sensitive underspecification and the acquisition of phonemic contrasts. *Journal of Child Language* **23**. 57–70.

Dinnsen, D. A. (1999). Some empirical and theoretical issues in disordered child phonology. In Ritchie, W. C. & Bhatia, T. K. (eds.), *Handbook of child language acquisition*. New York: Academic Press. 647–674.

Dinnsen, D. A. & McGarrity, L. W. (1999). Variation and emerging faithfulness in phonological acquisition. In Greenhill, A., Littlefield, H. & Tano, C. (eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. 172–183.

Dinnsen, D. A., McGarrity, L. W., O'Connor, K. M. & Swanson, K. A. B. (2000). On the role of sympathy in acquisition. *Language Acquisition* **8**. 321–361.

Dinnsen, D. A. & O'Connor, K. M. (2001). Typological predictions in developmental phonology. *Journal of Child Language* **28**. 597–618.

Eckman, F. R. & Iverson, G. K. (2000). Principles of L2 phonology: the question of learnability. In Juffs, A., Talpas, T. W., Mizera, G. & Burtt, B. (eds.), *University of Pittsburgh working papers in linguistics: proceedings of GASLA IV*. Pittsburgh, PA: University of Pittsburgh. 85–105.

Gierut, J. A., Morrisette, M. L. & Champion, A. H. (1999). Lexical constraints on phonological acquisition. *Journal of Child Language* **26**. 261–294.

Gnanadesikan, A. E. (1996). Child phonology in optimality theory: ranking markedness and faithfulness constraints. In Stringfellow, A., Cahana-Amitay, D., Hughes, E. & Zukowski, A. (eds.), *Proceedings of the 20th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. 237–248.

Goldrick, M. (2000). Turbid output representations and the unity of opacity. In Hirotani, M., Coetzee, A., Hall, N. & Kim, J.-Y. (eds.), *Proceedings of the North East Linguistic Society* **30**. Amherst, MA: GLSA. 231–245.

Hale, M. & Reiss, C. (1998). Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry* **29**. 656–683.

Hayes, B. (1999). Phonological acquisition in optimality theory: the early stages. Ms., University of California at Los Angeles. [Rutgers Optimality Archive #327].

Itô, J. & Mester, A. (1999). On the sources of opacity in OT: coda processes in German. Ms., University of California at Santa Cruz. [Rutgers Optimality Archive #347].

King, R. D. (1973). Rule insertion. *Language* **49**. 551–578.

Kiparsky, P. (1973). 'Elsewhere' in phonology. In Anderson, S. R. & Kiparsky, P. (eds.), *A Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston. 93–106.

Kiparsky, P. (2000). Opacity and cyclicity. *The Linguistic Review* **17**. 1–15.

Kirchner, R. (1996). Synchronic chain shifts in optimality theory. *Linguistic Inquiry* **27**. 341–351.

Macken, M. A. (1980). The child's lexical representation: the 'puzzle-puddle-pickle' evidence. *Journal of Linguistics* **16**. 1–17.

Major, R. C. (1987). Phonological similarity, markedness and rate of L2 acquisition. *Studies in Second Language Acquisition* **9**. 63–82.

McCarthy, J. (1999a). Morpheme structure constraints and paradigm occultation. In Gruber, C., Higgins, D., Olson, K. & Wysocki, T. (eds.), *Papers from the Regional Meeting of the Chicago Linguistic Society* **34**, vol. 2: *Papers from the panels: the status of constraints*, *the* *acquistion of spoken language, acquisition and the lexicon: April 17–19, 1998*. Chicago: Chicago Linguistic Society. 125–150.

McCarthy, J. (1999b). Sympathy and phonological opacity. *Phonology* **16**. 331–399.

McCarthy, J. (1999c). Sympathy, cumulativity and the Duke-of-York gambit. In Baertsch, K. & Dinnsen, D. A. (eds.), *Indiana University working papers in linguistics* **1***: Optimal green ideas in phonology*. Bloomington, IN: IULC. 57–91.

McCarthy, J. & Prince, A. (1995). Faithfulness and reduplicative identity. In Beckman, J. N., Dickey, L. W. & Urbanczyk, S. (eds.), *University of Massachusetts occasional papers* **18**: *Papers in optimality theory*. Amherst, MA: GLSA. 249–384.

Menn, L. (1983). Development of articulatory phonetic and phonological capabilities. In Butterworth, B. (ed.), *Language production* **2***: Development, writing and other language processes*. New York: Academic Press. 3–50.

Powell, T. W., Miccio, A. W., Elbert, M., Brasseur, J. A. & Strike-Roussos, C. (1999). Patterns of sound change in children with phonological disorders. *Clinical Linguistics & Phonetics* **13**. 163–182.

Prince, A. & Smolensky, P. (1993). *Optimality theory: constraint interaction in generative grammar* (*Technical Report* **2**). New Brunswick, NJ: Rutgers Center for Cognitive Science, Rutgers University.

Pullum, G. (1976). The Duke of York gambit. *Journal of Linguistics* **12**. 83–102.

Smith, N. V. (1973). *The acquisition of phonology: a case study*. Cambridge: Cambridge University Press.

Smolensky, P. (1993). Harmony, markedness and phonological activity. Paper presented at the Rutgers Optimality Workshop, New Brunswick, NJ.

Smolensky, P. (1996a). On the comprehension/production dilemma in child language. *Linguistic Inquiry* **27**. 720–731.

Smolensky, P. (1996b). *The initial state and 'richness of the base' in optimality theory* (*Technical Report JHU-CogSci*-96-4). Baltimore, MD: Department of Cognitive Science, Johns Hopkins University.

Stoel-Gammon, C. & Stemberger, J. P. (1994). Consonant harmony and phonological underspecification in child speech. In Yavas, M. (ed.), *First and second language phonology*. San Diego, CA: Singular. 63–80.

Tesar, B. & Smolensky, P. (1998). Learnability in optimality theory. *Linguistic Inquiry* **29**. 229–268.

Authors' address: Department of Linguistics, Indiana University, Memorial Hall 322, 1021 East 3rd Street, Bloomington, IN 47405, U.S.A.
E-mail: dinnsen@indiana.edu