Testing for OO-Faithfulness in the Acquisition of Consonant Clusters

Anne-Michelle Tessier

University of Alberta

Available online: 16 Apr 2012

To cite this article: Anne-Michelle Tessier (2012): Testing for OO-Faithfulness in the Acquisition of Consonant Clusters, Language Acquisition, 19:2, 144-173

To link to this article: http://dx.doi.org/10.1080/10489223.2012.660552

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Testing for OO-Faithfulness in the Acquisition of Consonant Clusters

Anne-Michelle Tessier

University of Alberta

This article provides experimental evidence for the claim in Hayes (2004) and McCarthy (1998) that language learners are biased to assume that morphological paradigms should be phonologically-uniform—that is, that derived words should retain all the phonological properties of their bases. The evidence comes from an artificial language word-learning paradigm, in which children learned novel objects names such as *wutch* and a plural suffix *-del* in an alien language and then were asked to produce alien words with difficult coda-onset clusters, some of which straddled the singular +*del* suffix boundary. The results suggest that 4-year-olds who are acquiring novel consonant clusters are preferentially faithful to the base segments in a derived word, e.g. *wutch* in plural *wutchdel*.

The article interprets these results from the perspective of Optimality Theory (Prince & Smolensky 2004; Tesar & Smolensky 2000), using Output-Output Faithfulness constraints (Benua 2000) to understand the asymmetries in the observed phonologies of derived and underived test items.

1. INTRODUCTION

1.1. Introduction to the Study

Much of the OT literature in phonological acquisition, in both the theoretical learnability and experimental domains, has addressed the role of learning biases. As in previous theoretical frameworks, learning biases all attempt to ensure that the learner obeys some version of the Subset Principle, adopting the most restrictive or conservative grammar consistent with the ambient data (Berwick 1985; Wexler & Manzini 1987; Dresher & Kaye 1990; Dresher 1999; in OT, see Smolensky 1996; Tesar & Smolensky 2000; J. L. Smith 2000; Prince & Tesar...
The basic learning bias in the OT literature is the ranking of Markedness \(\gg\) Faithfulness. Numerous empirical studies, starting with Jakobson’s (1941/1968) observations, have presented data with the M \(\gg\) F bias—in other words, evidence that children begin with an unmarked grammar and learn to produce increasingly more marked structures over time (e.g., Demuth 1995; Pater 1997; Levelt, Schiller & Levelt 1999; Kehoe 2000; Pater & Barlow 2003; Gnanadesikan 2004; Goad & Rose 2004; Levelt & van de Vijver 2004). Parallel to these observations, the M \(\gg\) F ranking bias has also been shown as necessary to preventing the acquisition of ‘superset grammars,’ i.e., incorrect grammar hypotheses about the target language, which positive evidence alone will never disprove (see especially Smolensky 1996).

This article reports an empirical investigation of another learning bias, proposed in Hayes (2004) and McCarthy (1998): the preference for high-ranking paradigm uniformity constraints, formalized here using Output-Output Faithfulness (OO-Faith; Benua 2000).\(^1\) OO-Faith constraints require that morphologically derived words retain the phonological properties of their morphological bases, and it has long been observed that target phonologies often impose phonological uniformity within their morphological paradigms (see the example in section 1.2, as well as examples in, e.g., Kenstowicz 1997, Steriade 1998, Benua 2000, and Pater 2000.) As Hayes (2004) and McCarthy (1998) have demonstrated, a bias for high-ranking OO-faith provides both a mechanism for preventing the potential acquisition of superset grammars (see section 1.2) and an explanation for anecdotal reports of learners who impose paradigm uniformity without evidence from the target language (see section 1.3). These latter examples give particular support to the notion that paradigm uniformity is a given in phonological learning: if children consistently innovate a pattern that does not approximate the ambient language, its motivation seems likely to reside instead within the acquisition process itself.

This study focuses on its predictions about phonological behavior at the point of morphological acquisition: i.e., when derived words are first decomposed into multiple morphemes and phonological patterns become attributable, in principle, to the demands of OO-Faith. Driven by these predictions, the study reported in section 2 used an artificial language game somewhat similar to a Wug-test (Berko 1958) to test 4-year-old children’s production of marked consonant clusters in two different morphological environments. The results in section 3 confirm that children often prefer to repair clusters in ways that satisfy OO-Faith at the expense of other pressure. Section 4 provides an OT interpretation of these results and the grammars that could derive them, as well as a discussion of this artificial language task. Section 5 discusses some implications and alternatives, and section 6 concludes the article.

1.2. The OO-Faith Ranking Bias

Two learnability arguments have been made for the high-ranking OO-Faith bias. One comes from McCarthy (1998), who points out that an inherent ranking of OO \(\gg\) IO-faith is necessary to learn a kind of OO-faithful language without surface alternations. In such a language, roots are unmarked in some static way, and this property holds of all members of a root’s morphological paradigm even when that markedness pressure is not present in the derived

\(^1\)Nothing crucial hinges on the choice of OO-Faith constraints to enforce paradigm uniformity; other constraint-based proposals of paradigm uniformity (see, e.g., Burzio 1996, Kenstowicz 1997, and Steriade 1998) would change the details of the analyses but not the force of the arguments. See however footnotes 3 and 4.
forms. In classical generative phonology, such effects were produced using Morpheme Structure Constraints (MSCs), which are untenable under the OT assumption of a rich base (Smolensky 1996; Prince & Smolensky 2004), but McCarthy (1998) shows how MSCs can be reinterpreted as the effect of OO-faithfulness. This work also points out that ensuring learners’ acquisition of the correct pattern for both simple and derived forms requires a bias for OO-faithfulness over IO-faithfulness, as the lack of alternations will not provide any overt evidence for their ranking.

The second learnability argument comes from Hayes (2004), who raises the issue of how children could acquire allophones whose normal phonological distribution is overridden just to keep a morphological paradigm uniform. The famous example he discusses is the interaction of flapping and Canadian raising (CR) in some dialects of English. In such dialects, CR is purely allophonic in monomorphemic words: raised [AI] appears before voiceless obstruents, as in write [Iart], while [ar] appears elsewhere as in ride [Iard]. However, derived forms with a base vowel [AI] exceptionally retain their raised quality even before a voiced flap, as in writer [Iairt]. [*] [Iairt]. Understood from the perspective of paradigm uniformity, writer has a raised vowel because it was derived from the base write, and in the base that raised vowel is phonologically-regular.2

The fact that morphological relations among words is necessarily learned after some basic phonotactics has been acquired puts any learner of this English dialect in a bind.3 At the very earliest stages, a learner without knowledge of underlying representations can acquire rough phonotactics by adopting the ‘Identity Map Hypothesis,’ i.e., assuming that observed outputs are identical to their target inputs, and so attempting to build a phonological grammar that will map any of these inputs faithfully onto themselves (see again Prince & Tesar 2004; Hayes 2004). At the point that such a learner encounters the word writer, how does he or she account for the presence of marked [AI] before a voiced flap? The child who does not yet know that writer is derived from write cannot explain this exceptional via morphological structure—instead it simply suggests that [AI] is a licit sequence of the target language, and indirectly that [AI] and [ar] are phonemic. The grammatical consequence of this supposed phonemic contrast is that IO faithfulness to the properties that distinguish [AI] and [ar] raised vowels (say, IDENT-[LO]) must rank above whatever markedness constraints disprefer raised vowels (simply referred to as *AI throughout this article) (see Table 1).

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>IDENT-[LO]</th>
<th>*AI</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>IAI</td>
<td>*AI</td>
</tr>
<tr>
<td>(b)</td>
<td>*IAI</td>
<td></td>
</tr>
</tbody>
</table>

Note. IDENT-[LO] >> *AI, given an unanalyzed input.

---

2In rule-based terms, the exceptional vowel allophone of writer is derived through rule ordering: first the diphthong is raised before /t/, then the stop is flapped and becomes voiced, erasing the surface environment for raising. Many if not most cases of paradigm uniformity can be derived by rule ordering and the related Phonological Cycle; see especially Brame (1974) and Mascaro (1976), and the Lexical Phonology of, e.g., Kiparsky (1982) and Mohanan (1986).

3See the summary of evidence for this chronology in Hayes (2004).
Unfortunately, the grammar in Table 1 is a superset of the target grammar, shown below in Table 2. In the correct ranking, \(^{\star}\text{AI}\) is ranked above IDENT[LO]-IO, and the exceptional raised vowel in writer is possible only because of its required similarity to write, derived using IDENT[LO]-OO. Notice that compared to Table 1, not only has the ranking in Table 2 changed but so too has the input: the underlying form of writer has been broken down into two morphemes, and write has been identified as the base to which derived forms should be faithful (see Table 2).

Hayes (2004) shows that a learning bias for high-ranking OO-Faith offers an answer to the question of how learners can eventually reach the grammar in Table 2, despite their initial error in Table 1, by providing a pressure that will return the grammar to a more phonologically restricted state once morphological learning has occurred. This bias places the set of OO-faithfulness constraints in the initial state at the very top of the grammar, as in (1):

\[
\text{(1) Initial state ranking} \ni \\
\text{OO-faithfulness} \gg \text{Markedness} \gg \text{IO-faithfulness}
\]

Before the morphology of words like writer is known, the learner will still incorrectly attribute the exceptional allophone to IO-faith, as in Table 1, with OO-faith vacuously satisfied (see Table 3).

When the learner realizes that writer includes the base write, however, its UR can be revised to match the two morpheme input in Table 2 above. With this shift, the high-ranking OO-faith constraint can now choose the winner also as in Table 2, and return the learner to the correct final-state ranking.\(^4\)

\(^4\)This view of paradigm uniformity relies crucially on a base and on the notion of ‘base priority’ as in Benua (2000)—that is, it is not compatible with the Optimal Paradigms approach (McCarthy 2002a) in which every member of an inflectional paradigm influences every other member. For arguments that inflectional paradigms do in fact have bases, see Albright (2008).
1.3. Innovative OO-Faith in Early Phonology

1.3.1. OO-Faith in the Anecdotal Literature

The sketch above of Hayes’s (2004) example shows how an OO-faith bias helps the learner to acquire morphologically sensitive phonology if they are present in a target language. But this bias also predicts innovative morphological sensitivities—that is, overly OO-faithful errors that learners should make in the acquisition of target phonology that shows no sensitivity to basehood in the adult grammar. In fact several such cases have been noted in the literature, in which a child appears to temporarily override phonological regularities just in derived words. Hayes (2004) reinterprets one such pattern described by Kazazis (1969), of one Greek-learning child who at 4;07 correctly produced velar and palatal fricative allophony except where it disrupted the paradigm uniformity of verbal stems. Hayes also points to Bernhardt and Stemberger’s (1998) report of a child who from 2:00—3:08 consistently flapped in simplex words (e.g. water [‘wa.ri]), but only produced base-faithful voiced and voiceless stops in derived words, even where the adult phonology requires a flap (e.g. sitting [‘sttr] from base [str] and ‘needed’ [‘nid] from base [nid]). A third case comes from the famous puzzle- puddle-pickle chain shift reported by N. V. Smith (1973) and further analyzed by Macken (1980). In this child’s grammar, coronal stops generally became velar before liquids (puddle → puggle), but this pattern was blocked when the coronal stop formed part of a morphological base (tightly, *tigh[k]ly, with base tigh[t])—on this point see especially Dinnsen & McGarrity (2004) and Jesney (2005). An interesting additional case comes from the acquisition of Hebrew verbal morpho-phonology, analyzed in somewhat different terms by Adam & Bat-El (2008). This study reports that children learning Hebrew systematically produce bare verbs without inflectional suffixes—a child-specific innovation, given the rarity of unsuffixed forms in the ambient language—even when children’s noun production indicates that their phonology can tolerate the prosodic shape of suffixed verbs. In many cases this omission of suffixes can be understood as a result of OO-faithfulness to the verbal stem: e.g., since many suffixes are stress-shifting, omitting a suffix means maintaining stress in its base location (on such OO-faith constraints against stress shift, see Kager 1999; see Gouskova 2007 for the related point that DEP-OO can minimize affixation in derived forms.)

The hallmark of all these attested child patterns is that the same structure is treated more faithfully when affiliated with a morphological base rather than otherwise, and that this preferential faithfulness is not supported by any target language evidence. To clarify these effects, the rest of this section shows how high-ranking OO-faith constraints can create paradigm uniformity effects at three stages of acquisition and so illustrates the patterns that are predicted to emerge in the experiment to follow if the OO-faith bias obtains.

1.3.2. OO-Faith Stages Schematized

The markedness constraint used in all these examples is AGREE[VOICE] (Lombardi 1996, 1999), which requires obstruent clusters to match for voicing, either all voiced or all voiceless. The conflict between faithfulness constraints and AGREE[VOICE] can be tracked using two lexical items of the experimental language to come, given in (2):

See Bernhardt & Stemberger (1998) for their alternative analysis of this data.
This choice of example is not to say that voicing agreement is the only or even the primary error seen in the experimental results, but it will serve only as one example in this section.

At the initial state from Table 3, learners should respect all markedness constraints except at the expense of violating paradigm uniformity. Before morphological learning, this exception clause is vacuous since both inputs are unanalyzed wholes, /\textipa{\textipa{watfd\textipa{d}}}/ and /\textipa{\textipa{zrfdtn}}/, and so AGREE[VOICE] causes voicing assimilation in both words, in whichever way the set of IO-faith constraints deems most harmonic. The evidence from both child and adult grammars is that when possible, onsets are usually preserved more than codas (see, e.g., Lombardi 1999; Chambless 2006); to capture this asymmetry, this article assumes two kinds of IO-faith constraints, namely general IO-faith constraints and more specific IO-faith-Onset constraints (Beckman 1998). In the present example, then, the grammar includes both IDENT-[VOICE] and IDENT-[VOICE]-ONSET. With these IO-faith constraints, the grammar will always repair voicing mismatches by changing input coda voicing to match the onset\(^6\) (see Table 4).

How does the initial stage grammar treat these two words after morphology is learned? Voicing mismatches will still be repaired, but repaired differently depending on whether the cluster spans a base/affix boundary or not. In simple forms, the repair remains the same as in Table 4: voicing is protected in onset, so it is changed in coda, and OO-faith has no role to play. In derived forms, however, coda voicing is now protected by OO-faith as part of a morphological base, so voicing will change in onset (see Table 5).

To give a second example, a similar repair asymmetry will be produced at this postmorphological initial state by the conflict between a constraint on coda affricates and MAX (the basic constraint prohibiting deletion)\(^7\) (see Tables 6 and 7).

A third nontarget grammar is beyond the initial state, where some constraints have been reranked, but full mastery has yet to be achieved. The present goal is merely to diagnose the possible effects of these constraints and their ranking; the question of how and why learners progress from the initial state to any others is left aside here until section 5.2. But whatever

\(^6\)In Table 4, the ranking of the specific and general IO-faith constraints is undetermined. It has been argued that a learner should in fact be biased for the former such constraints to outrank the latter, but in this article such a ranking is not crucial; see J. L. Smith (2000), Hayes (2004), and Tessier (2007, 2009); cf. Prince and Tesar (2004).

\(^7\)Ensuring that the definition of an onset-specific MAX constraint can penalize this deletion of [d] without syllabifying the input—see McCarthy (2002b) for reasons behind this assumption—is nontrivial. For reasons of space I will simply assume here that it can be done; cf. Beckman (1998).
the mechanism of learning, the final state grammar that tolerates voicing mismatch in both morphological contexts is one in which markedness ranks below general IO-faithfulness. Along the way to this final state—with just these four constraints and two-word lexicon—there is only one ranking that is more faithful than the initial state in Table 4 but is not yet fully faithful as the target. This ranking is one in which markedness, \textit{AGREE-VOICE}, has been demoted below faith to onset voicing alone, but the other two faith constraints remain above (their respective ranking being noncrucial):\textsuperscript{8}

\begin{equation}
\text{(3) Intermediate stage ranking} \\
\text{Id-[VOICE]-OO, Id-[VOICE]-ONS-IO} \gg \text{AGREE-VOICE} \gg \text{Id[VOICE]-IO}
\end{equation}

Like the last stage, this grammar also derives asymmetric results in the two morphological contexts. As before, underived forms show the familiar change in coda voicing (see Table 8).

\textsuperscript{8}For evidence of such stages in natural language acquisition characterized in these terms, see Tessier (2009) and references therein.
In derived forms, however, the effect of this intermediate reranking is now to block all repairs for cluster voicing mismatches: OO-faith blocks voicing change in coda, and onset-specific IO-faith blocks a repair in onset. As a result, the optimal candidate is one in which voicing disagrees, violating midranking markedness (see Table 9).

To summarize these asymmetries,

(i) at initial state, the location of unfaithfulness should be influenced by basehood, such that derived forms alone will show exceptional repairs, which affect non-base segments; and
(ii) at an intermediate stage, the degree of unfaithfulness should be sensitive to morphological bases, such that derived forms will be more faithful than underived ones.

Of course, the existence of these asymmetries in learning is created through the confluence of several factors: the position of the morpheme boundary (between coda and onset), a markedness constraint that can be repaired on either side of that morphological boundary, and a positional faith constraint to onsets alone (where the marked material that does not belong to the base appears.) But this careful cluster of properties is intentional: the experiment to be reported was designed to create just this confluence, and the data were examined by comparing the same strings (like /tʃd/) in these two morphological environments.

1.4. Research Questions

The primary question of this study was whether children demonstrate any OO-faith preferences, independent of target language evidence, in a controlled experimental context. Section 2 below lays out the rationale and specifics of the study designed to probe this question, and section 2.5 operationalizes this goal as a pair of experimental predictions given its design. The secondary question, which the former relies on crucially, was to whether 4-year-old children are able and
willing to engage in artificial language wug testing at all—whether an hour-long study session would be sufficient for them to accept a novel plural suffix and use it productively.

2. THE EXPERIMENTAL STUDY

2.1. A Methodology to Test for an OO-Faith Bias

Investigating the existence of innovative paradigm uniformity in natural L1 phonological learning poses some large challenges. For one thing, the predicted effects of OO-faith at any stage are intrinsically tied to the learner’s representational assumptions about morphology—its bases, relations, paradigms, and the like—and children’s acquisition and interpretation of morphology may vary considerably. Furthermore, English morphology does not provide many good testing grounds for such investigation.

The approach of this study was to use an artificial language learning study, in which 4-year-old children learned both unfamiliar coda-onset clusters and a novel plural suffix in an explicitly ‘alien’ language invented by the experimenter and taught to children by an alien puppet named Bozdim, operated by the experimenter. This design was an attempt to create a stage at which a piece of morphology had been acquired (the novel plural suffix /-daʃ/) while some phonology had not (e.g., the marked coda-onset clusters /tfd/), so that the effects of an OO-faith bias might be detected. Over the course of the experiment, a type of word-learn ing paradigm was used to probe for asymmetries in the phonology of simple vs. complex words.

2.2. Participants

Twelve 4-year-old children (mean age 4;07) in the Amherst and Northampton, Massachusetts, areas learning English in monolingual homes participated in the study. This age group was chosen because 4-year-olds typically have fairly advanced phonological grammars, but have not completely mastered difficult and/or infrequent segments and clusters. With respect to morphology, 4-year-olds are also reported to be at the stage of overgeneralizing regular morphological patterns—e.g., foots, mouses; ranned, bringed—and thus presumably in the throes of productive morphological acquisition.

2.3. Materials

Table 10 below lists the novel words children were taught in experiment, organized by their relevant properties. Count noun singulars were all monosyllables of the shape CVC(C); each singular was suffixed with [-daʃ] to form a bisyllabic plural, CVC(C)daʃ. Mass nouns were all bisyllabic with the shape CVC(C)VC. All bisyllabic forms had stress on the initial syllable; all vowels in the second syllable were lax (ı̈, e, or a) and unstressed but not completely reduced.

---

9 The study was piloted with 3-year-olds, but it was quickly determined that they were uninterested or unable to engage in the task, and in particular did not seem to understand the concept of the alien puppet speaking an ‘alien’ language sufficiently to imitate him with communicative intent.

10 In fact, two of the children whose data is reported here produced English morphological errors of this sort during the spontaneous conversation of the experiment: “He brigned the chair” and “I runned to the table.”
Crucially, every cluster occurred both within a mass noun and across the count noun-plural suffix boundary\textsuperscript{11} (see Table 10).

The clusters used were intended to represent both uncommon and impossible English coda-onset clusters. (Initial attempts to pilot the experiment showed that only using impossible English clusters resulted in participant frustration and quick drop-outs, while only uncommon clusters produced too few errors to analyze.) To quantify, the type frequency of these word-internal clusters in the Carnegie Mellon University (CMU) pronouncing dictionary (http://www.speech.cs.cmu.edu/cgi-bin/cmudict, version downloaded June 27, 2010) range between 55 (for \[b.d\]) and 0 (for \[mf.d\]) out of a total 133,700 types, representing a fairly vanishing number of such clusters.

The mass noun stimuli were created to be as similar phonologically to the plural nouns as possible, while keeping them distinct from the plurals.\textsuperscript{12} While the first syllable codas were held constant across the two conditions, Table 10 reveals that the vowel qualities of the first syllable in plural vs. mass nouns were not. As a reviewer points out, trimoraic syllables are not found word-internally in English (although they are attested cross-linguistically in, e.g., Finnish; see Suomi, Tovanen & Ylitalo 2008). In part this was deliberate, as it was hoped that overly long syllables like \[\text{ka}\text{b}\text{d}\text{d}\] would emphasize to participants the non-native phonology of the language being learned. However, a post-hoc determination of the cumulative probability of each plural and mass noun was calculated,\textsuperscript{13} and both morphological

\begin{table}[h]
\centering
\caption{Full Set of Materials Used in the Study}
\begin{tabular}{|c|c|c|c|c|}
\hline
\textbf{Coda Segment(s)} & \textbf{Cluster} & \textbf{Count Singular} & \textbf{Count Plural} & \textbf{Mass} \\
\hline
\textbf{Stop(s)} & bd & pob & pobd\textsuperscript{\textdagger} & gribd\textsuperscript{\textdagger} \\
& gd & wag & w\textsuperscript{\textdagger}g\textsuperscript{\textdagger}d & mogdem \\
& kd & p\textae{k} & p\textae{k}\textsuperscript{\textdagger} & l\textae{k}t\textae{d}\textae{n} \\
& ktd & l\\textae{k}t & l\textae{k}\textae{t}\textae{d}\textae{a}\textae{t} & paekt\textae{d}r\textae{m} \\
\textbf{Fricative-Stop} & f\textae{dt} & faeft & faeft\textsuperscript{\textdagger} & l\textae{f}t\textae{d}\textae{sk} \\
\textbf{Affricate} & tf\textae{d} & w\textae{t}\textae{\textae{f}} & w\textae{t}\textae{t}\textae{f}\textae{d}\textae{a}\textae{t} & z\textae{t}\textae{t}\textae{d}\textae{n} \\
& d\textae{zd} & b\textae{rd}\textae{\textae{z}} & b\textae{r}\textae{d}\textae{z}\textae{d}\textae{a}\textae{t} & f\textae{d}\textae{g}\textae{d}\textae{t} \\
& b\textae{zd} & g\textae{r}\textae{b}\textae{z} & g\textae{r}\textae{b}\textae{z}\textae{d}\textae{a}\textae{t} & m\textae{e}\textae{b}\textae{g}\textae{d}\textae{t} \\
\textbf{Nasal-Fricative} & m\textae{fd} & naem\textae{f} & naem\textae{f}\textae{d}\textae{a}\textae{t} & g\textae{m}\textae{f}\textae{d}\textae{p} \\
\hline
\end{tabular}
\end{table}

\textsuperscript{11}A reviewer raises the question of whether more variability in prosodic shape would have been beneficial to the study, for example including bisyllabic singulars or monosyllabic mass nouns. Two reasons to keep the materials as short as possible: first, to limit the already significant memory demands on participants, and second, to keep the stress profile of singulars and plurals constant across all items.

\textsuperscript{12}A reviewer suggests that the fact that the lack of variability in the phonological shape of mass nouns—e.g., that their second syllables was always –dVC—might have caused participants to have learned a second templatic affix marking mass nouns, but nothing in the data indicates such a misinterpretation.

\textsuperscript{13}The cumulative probability of each word was determined by summing the natural log likelihood of each of its biphone probabilities, calculated across the entire CMU pronouncing dictionary; using such a measure to determine word acceptability in particular is supported by results in, e.g., Coleman & Pierrehumbert (1997) and Frisch, Large & Pisoni (2000).
categories emerged as comparably improbable (for plurals, the mean and standard deviation of the items cumulative probabilities was $-23.72 (4.00)$, while for mass nouns it was $-26.38 (5.06)$.)

As explained further below, participants learned these alien words by hearing them spoken “by the puppet”—that is, spoken by the experimenter playing the part of the puppet. This design was chosen over one using recorded materials to allow for as spontaneous discourse as possible between participant and puppet, but it does however necessarily raise the question of whether the experimenter’s model productions might have influenced their acquisition. Every effort was made by the experimenter to produce the clusters and their segments similarly in all tokens and contexts: in particular, coda stops in all contexts were slightly released.\textsuperscript{14} These measures were taken to make mass and count plural nouns as prosodically similar as possible, but to maximize the perceptability between the final rimes of the plurals (always [-d\textsubscript{\textcircled{r}}]) and the mass nouns (always of the form [dVC].)

One potential confounding difference raised by a reviewer is that plurals nouns could have been produced in a way that cued a two-word interpretation—e.g., with less consonantal overlap and longer overall syllable timing—as a by-product of the experimenter’s attempt to produce each morpheme clearly. On the other hand, another reviewer suggests that mass nouns with trimoraic syllables might have been interpreted as compounds or otherwise poly-morphemic on phonological grounds alone. As one test of these hypotheses, average lengths of mass and count plural nouns produced by the experimenter were computed for all 12 sessions, comparing the lengths of the entire tokens as well as each of the syllables. On average, the plural tokens were slightly longer than the mass tokens produced by the experimenter, and in fact the second syllable of the mass nouns was slightly longer than the second syllable (i.e., the suffix) of the count plurals, but neither comparison proved significant.\textsuperscript{15} It did however prove the case that the experimenter produced the first syllables of plural nouns with an average length of 374 msc, whereas the first syllables of mass nouns were produced with an average length of 357 msc, and that this difference bordered on significant ($t(130) = 1.95$, $p = .053$). It cannot be entirely ruled out that this 17 millisecond difference in duration caused children to interpret the root syllable of the plural tokens as constituting the first half of a plural compound, but since the second syllables were if anything shorter in plurals compared to mass nouns it is not clear that this durational cue was supported throughout the word.

\subsection*{2.4. Methods}

Initially the experimenter and puppet, Bozdim, presented the child with pictures of familiar objects one at a time. The child was asked for the English name for the object and then engaged in a short discussion of the object and its properties—its color, size, prototypicality, etc.—to get the child focused on the object. Then, the puppet was asked to give the name of the object in his language,\textsuperscript{16} and the experimenter (playing the part of the puppet) would produce the

\textsuperscript{14}As a reviewer points out, this made the word-internal coda consonants sound rather foreign (particularly the voiced ones) as such English consonants are not typically released.

\textsuperscript{15}Entire tokens: average plurals 665 msc; average mass 637 msc, $t(130) = 0.834$, $p = .405$. For syllable 2: plurals 274 mscs, mass 294 mscs, $t(130) = 1.262$, $p = .209$.

\textsuperscript{16}Ideally, the child would ask the puppet, but in the face of shyness the experimenter would do so instead.
Once the puppet had given the object’s name, the child was encouraged to repeat the name (“Can you say what Bozdim just said?”) and to use it in similar discussion as with the English name before (“Is this a blue wutch? Or is it a yellow wutch?” or “Does it look like this cup of zitchdin tastes good? Do you think the zitchdin is hot or cold?”)

In this phase, children learned six object names, in two blocks of three. Within a count noun block, participants first learned three singular nouns and then their corresponding plurals. Within a mass noun block, participants learned three mass nouns and then saw them again in a different container—e.g., a glass of juice and, later, a bottle of juice. The order of count vs. mass noun blocks was counterbalanced across participants. A sample set of materials that children learned in this initial block is given below (each child learned the words in a random order—except that the easiest singular *pob* was always presented first) (see Table 11).

After the initial block, the experimenter asked the children to play a matching game with the puppet. All 12 pictures seen so far were laid out in front of the child; to play the game, the puppet pointed to one picture and named it for the child, and the child would then point to the matching picture and name it for the puppet. In the game, the puppet pointed to one of each of the mass nouns for the child to match by naming the other, and to each of the singular count nouns for the child to match by naming the plural. Thus, this matching game required the child to provide six bisyllabic words: three underived mass nouns and three plural nouns derived from singular bases (these six words are in bold in the sample block above.) This game was therefore a rather simplified version of the Wug-test: while children were never asked to produce a plural for a count noun they had only ever learned in the singular, the context of the matching game had only recently provided the singular form, whereas the plural would have been heard in the earlier experimental phase.

After the first game, children were presented another training block, as in Table 11 below, with the remaining items, followed by testing of the new words. This second block included either two or three more words of each category, depending on the engagement of the child.

<table>
<thead>
<tr>
<th>Noun class</th>
<th>Morphology</th>
<th>Prosodic shape</th>
<th>Words</th>
<th>Sample matching pictures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>Three singulars (base)</td>
<td>CVC(C)</td>
<td>pob</td>
<td>one armchair</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>watf</em></td>
<td>one pick-up truck</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>naemf</td>
<td>one flower</td>
</tr>
<tr>
<td></td>
<td>Three plurals (base + suffix)</td>
<td>CVC(C).də+</td>
<td><em>pobdə+</em></td>
<td>many armchairs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>watdə+</em></td>
<td>a fleet of pick-up trucks</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>naemfrət</td>
<td>a garden of various flowers</td>
</tr>
<tr>
<td>Mass</td>
<td>Three mass nouns</td>
<td>CVC(C).dVC</td>
<td><em>gəbðrət</em></td>
<td>a glass of juice</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>zətfdən</em></td>
<td>a cup of hot chocolate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>gəmfdəp</em></td>
<td>a mug of milk</td>
</tr>
<tr>
<td></td>
<td>Same three mass nouns</td>
<td>CVC(C).dVC</td>
<td><em>gəbðrət</em></td>
<td>a bottle of juice</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>zətfdən</em></td>
<td>many cups of hot chocolate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>gəmfdəp</em></td>
<td>a carton of milk</td>
</tr>
</tbody>
</table>
2.5. Predictions

The two specific predictions aimed at finding support for an OO-faith bias in this experimental context both relate to the treatment of the medial clusters in mass vs. plural nouns, once a morphological paradigm has been established:

Prediction 1: The first syllable coda (or coda cluster) of plural nouns should be equally or more faithfully preserved compared to the first syllable coda(s) of mass nouns.

Prediction 2: Unfaithfulness to the onset, rather than coda, of a medial cluster should be found only in plural nouns.

Prediction 1 is derived from the location asymmetry seen in section 1.3: depending on the ranking of constraints, codas might be faithful in both morphological environments, or in neither, or only in plurals, but no ranking prefers faithfulness only in mass nouns. Prediction 2 recapitulates the degree of faithfulness asymmetry—in mass nouns there is no pressure that does not prefer repairing in coda, whereas in plurals an undominated OO-faith constraint can drive repairs in onset position (as in Table 5).

3. RESULTS

3.1. Data Transcription and Coding

Each session was tape-recorded using a DAT tape recorder and table microphone; sessions tapes were then digitized as .wav files. The sessions were transcribed both by the experimenter and by a graduate student skilled in transcription and unaware of the goals of the experiment. Initially, between-transcriber agreement was 92%; for tokens where the transcribers disagreed, a consensus was reached, and the three tokens on which a consensus could not be reached were thrown out.

It should be noted before continuing, however, that the transcription of early child phonology can be complicated in a few ways, notably by the existence of “covert” contrasts (Scobbie et al. 2000), whereby a child may reliably produce a measurable phonetic difference between, e.g., target voiced and voiceless segments, but one that is either too subtle for adult transcribers to detect or else somehow misaligned with the native language boundary for VOT values in a way that obscures the distinction for transcribers. It has similarly been shown in some studies (Carter & Gerken 2004) that segments that appear to have been deleted in children’s productions can nevertheless leave a small durational trace in the acoustic signal. Since many of the errors reported in this section include either voicing changes or segmental deletions, these transcription concerns must be confronted.

With respect to the transcription of voicing, two criteria were used to determine that a production was coded as having a voicing error. First, if either transcriber felt a segment was faithful to the original voicing it was treated as such. If both thought its voicing was nontarget then the spectrogram was examined; if the spectrogram including a voicing bar in more than 50% of the stop closure or frication duration, it was considered voiced, and otherwise was

---

I am grateful to an anonymous reviewer for pushing this point.
considered voiceless. In cases where both transcribers thought there was unfaithfulness to voicing, this second criterion always corresponded with the transcribed voicing—that is, the spectrogram always confirmed the transcribers’ impression of nontarget voicing. These criteria were used for all children, and while they may have led to underdiagnosis of unfaithful voicing, they were at least held constant. The approach to segmental deletion was less systematic: the spectrograms for any purported deletions (as heard by both transcribers) were examined for any formant warping or silence between the perceived flanking segments, or any other residue of the deleted segment, but no easy comparison was possible between these tokens and those without a missing segment. However, even if these deleted segments were in fact only “mostly deleted,” these productions should still count as errors from this study’s perspective, in contrast to those tokens where all target segments were included. The much larger question of how such contrasts, when present, should be understood or represented must remain beyond the scope of the present work.

In addition, this experimental design’s reliance on morphological “awareness” requires careful decisions as to the data included: to test any of the predictions, we must be able to claim that participants had in fact learned the artificial language’s plural suffix. To prove sufficient mastery of –del, participants had to provide at least one spontaneous token of more than one plural noun, associated with the right plural picture.\(^\text{18}\) This criterion eliminated two participants, leaving 10 children whose data are discussed in the remainder of the article.

Of the nine clusters tested, only four are included in the final results. Two criteria excluded the others: first, the cluster had to be pronounced unfaithfully in more than two tokens by any one child; second, it had to have been produced by more than one child in both the mass and plural contexts. The first criterion eliminated two clusters (gd, kd), which were each only mispronounced once in the entire corpus of data; the second criterion ruled out another three clusters (ftd, ktd, b3d) which the participants simply refused to produce (the first two of these clusters were only attempted once by one child each; the latter was never produced at all.) Thus the remaining clusters for analysis were as in (4):

\[(4) \text{Final target clusters: } b,d \ mf,d \ tf,d \ d3,d\]

All results reported are for these 4 clusters and 10 children. All tokens are reported from what was referred to above as both training and testing—this compromise was adopted simply to get enough tokens.\(^\text{19}\) However, plural tokens were only included when the participant produced a second syllable of type dV(C), ruling out tokens with English plural affixes (two wutch[\text{az}], pob[\text{z}]) or zero morphology (two wutch, pob).

\(^\text{18}\)To be considered a “spontaneous” utterance, a token had to be produced before the experimenter did so, on the particular viewing of the picture. Thus the spontaneous utterances were almost all in the “testing” portion of the experiment.

\(^\text{19}\)A reviewer rightly points out that learning may well have taken place over the course of the experiment, whether or not during “training” or “testing.” While it seems unlikely that the children’s ranking of phonotactic constraints could be altered merely from fifty minutes exposure to these clusters—though we cannot be sure it was not—it does seem very reasonable that participants had to hear plural words enough times to correctly analyze their morphological structure and build inputs that allow OO-faith constraints to assess violations, as in the shift from Table 1 to Table 2 seen in section 1.2. Not knowing when this might have occurred for each individual participant is an unfortunate shortcoming of the methodology, though at least every child came to the study with an equal previous ignorance of the suffix.
3.2. Overview of the Results

Table 12 above gives a first indication of how much data each child provided during the course of the experiment, giving the total number of tokens produced in each of the three morphological environments. Across the 10 children, there were 107 singular count noun tokens, 87 count noun plurals, and 112 mass nouns (see Table 12).

The greater number of singulars compared to plurals may plausibly suggest that the plurals were sufficiently harder to sometimes be avoided; since the mass nouns were presented in two different contexts there was double the opportunity to elicit them, so the fact that they were only somewhat more numerous than the plurals suggests that they too were found difficult.

Of the 107 monosyllabic singulars, the overwhelming majority (104/107) were faithfully produced. Thus we have evidence that the CVC and CVCC structures in these bases were within the phonological reach of the participants and that their mispronunciation in plurals was not the result either of inventory gaps or incorrect lexical representations.

The majority of children’s two syllable words fell into two categories: either faithful or unfaithful in the coda of the first syllable. Such unfaithfulness comprised featural changes or deletion. The proportion of unfaithfulness productions by location and cluster is tabulated in Table 13.

---

20The entire corpus of data can be accessed online at http://www.ualberta.ca/~annemich/OO-FaithMaterials/
The second and third rows of the table, which together include all errors in the medial cluster, are the ones that will feature prominently in the rest of the discussion—before moving on, the remaining three categories will be dealt with briefly. The category ‘syllable boundary’ covers two types of errors: the metathesis of syllable 1’s coda and syllable 2’s onset and epenthesis between the two (e.g. [grbdrdt]). The errors in the initial syllable onset were usually a result of long distance place harmony (e.g., /grbdrt/ → [dibrdrt]), and the errors in the final coda were mostly place changes as well (/gamfd̩pt/ → [gamfd̩pt]; /z̩t̩∫f̩dm/ → [z̩t̩∫f̩dm]). It should be noted that only 1 of these 17 word-final errors included a mass noun’s second syllable replaced by ‘-del,’ so these errors do not represent a significant confusion between mass and plural final syllables.

3.3. Testing Predictions: Faithfulness to Medial Clusters

This section focuses more carefully on the treatment of medial clusters in the bisyllabic items, i.e., plurals and mass nouns. To begin, Table 14 below provides a breakdown of the kinds of errors recorded in these clusters.

The main prediction to be tested against the medial cluster data is that initial syllable codas should be either equally or more faithful in count nouns plurals compared to mass ones, as the plural codas are protected by membership in a morphological base. To examine this, Table 15 below shows the proportion of all tokens whose codas were faithful, produced by each child in the two morphological conditions.

There is a limit to the amount of confidence to be placed in any analysis of the small amount of data in Table 15, but two avenues of investigation are pursued here. First, the data from all 10 children can be pooled for an overall comparison of the mean coda faithfulness among plurals vs. mass nouns, as suggested by Prediction 1. Summing across participants, a two-tailed t-test

<table>
<thead>
<tr>
<th>Error Location</th>
<th>Type of Error</th>
<th>Cluster</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>b.d</td>
<td>dZ.d</td>
</tr>
<tr>
<td>σ1 Coda</td>
<td>voicing change</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>segment deletion</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>/C/ → [?]</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>/b/ → [v]</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>affricate → fricative</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>place change</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>other featural change</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>totals</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>σ2 Onset</td>
<td>voicing change</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>segmental deletion</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>affricate → fricative</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>totals</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
assuming unequal variance confirms that codas were produced faithfully significantly less often in mass noun clusters than in the plural count noun clusters ($t(18) = 3.754$, $p < .01$). Further, a paired two-tailed $t$-test comparing the proportions for each child also shows a significant difference between the lower proportions of faithful first-syllable codas produced in mass nouns compared to the higher proportions in plurals ($t(9) = 4.73$, $p < .01$). These positive results provide evidence for the kind of intermediate stage grammar seen in Tables 8 and 9 of section 1.3, where plural nouns are faithful to both members of medial clusters but mass nouns are still unfaithful in coda position.

The other approach to the data in Table 15 is a descriptive characterization of individual results; from this perspective, the table’s participants can be grouped into three broad categories. The first three children (I, A1, D2) show unfaithfulness in both morphological environments, while the second three children are roughly faithful to most codas, regardless of morphology (D1, C, N2). The third group, however, show a marked difference between morphological environments: for E, A2, A3, and N1, initial syllable codas for mass nouns are considerably more prone to unfaithfulness than those codas in plural nouns—these four participants together produced 12/36 faithful medial codas in mass nouns but a full 25/27 faithful medial codas in plurals.\textsuperscript{22} It is worth noting that in the first two groups, no child was less faithful in mass nouns than plurals in terms of token numbers in Table 15, though the data is sparse enough that nothing crucial should be extrapolated from this asymmetry.

\textsuperscript{21}Noting that the low number of total items for D1 and N1 might have skewed this result, this pairwise comparison is also significant for the first eight children alone.

\textsuperscript{22}N1 produced so few plurals that his data might be ignored, without reversing this trend.
The other prediction about medial clusters made in section 2.5 concerned the location of faithfulness—namely that unfaithfulness in a syllable 2 onset should be found only in a plural noun, where it could be driven by the need to not be unfaithful in the preceding coda from the singular base. This prediction is rather difficult to test, since only 11 onset repairs were observed in the entire data set. All 11 tokens are provided in Table 16, showing that 8/11 did in fact occur in plural nouns.

The existence of this asymmetry, such as it is, is important in part because it cannot be understood under an alternative memory-based interpretation of the plural vs. mass asymmetry. This account would suggest that plurals overall are easier than mass nouns because the former are made up of two one-syllable parts, the root and the plural suffix, and that two simple (e.g., monosyllabic) things are easier to remember than one complex (bisyllabic) thing. While this notion might explain a difference in degree of faithfulness (as in the intermediate stage asymmetry), it cannot explain any difference in the location of faithfulness, where the difficulty is not how much is remembered (or produced) correctly, but where the mismatch between correct and incorrect forms occurs.

### 3.4 Summary of Results

The results of the previous section were overall positive: the data from 10 children does show the influence of morphological structure on phonological faithfulness as predicted by high-ranking OO-faith, in terms of the developmental stages discussed in section 1.3. The next section turns to the grammatical analysis of these results, from the perspective of these proposed stages.

### 4 Modeling the Treatment of Medial Clusters

#### 4.1 L1 Transfer and the Nature of the Task

Adopting an artificial language learning paradigm requires at least one complicated set of

---

23 Thanks to anonymous reviewers for raising this point.
assumptions: those about the state and source of the grammar that participants are using in the task. So before presenting grammatical interpretations of Table 15’s results, this section discusses and justifies at some length the assumptions being made in these interpretations regarding the grammar that children might have used in this experiment and why.

It is clearly the case that participants do not come to this task at the ‘initial state’ in any natural sense: 4-year-old children have both an English morphological representation of ‘plural’ as well as an English-specific phonology of consonant clusters like those used in section 1.3’s tables. When facing explicitly foreign-phonology (e.g., words being translated by an acknowledged alien that contain at best unfamiliar clusters), what might a learner do?

I first assume that learners treat this data as evidence of a new, non-English phonology to be acquired, including the potential for morphological interactions with these sound patterns. Given the specificity with which phonological patterns can be sensitive to morphology, each new affix that the learner encounters should come with a blank-slate set of assumptions about its interaction with phonology. From the present perspective, this means that learning a new suffix like -del means creating a new del-specific OO-faith constraint and installing it as per the ranking bias at the top of the current phonological grammar.

As for the rest of this grammar below the OO-del-Faith constraints, two main options present themselves. One logical possibility is that children might construct a phonology from scratch for this new language, though this may seem implausible with only 45 or 50 minutes of exposure. More likely is that learners would transfer their current L1 phonology, as found for adult L2 phonologies (e.g., Schwartz & Sprouse 1996; Hancin-Bhatt & Bhatt 1997; also Broselow 2004); for recent evidence of L1 transfer in early L2 phonology, see Sorenson-Duncan (2010) and references therein).

While this study’s design cannot determine the extent to which participants were actively constructing or using an L1-transfered L2 phonology, I suggest that any such hypothesis should still allow us to test for OO-faith biases with this methodology. What the study did was probe the morphological influence on marked clusters that are, for the most part, not allowed in the L1—certainly the frequency of all the reported medial clusters apart from [b,d] in a typical 4-year-old’s experience should be akin to zero, and even experience of [b,d] should be very limited. What should transfer from the children’s English, then, would be a grammar that does not permit full faithfulness to these clusters, ranked below a novel set of OO-del-faith constraints.

One potential criticism of this view is that the children’s English phonology might already dictate that some marked properties of the medial clusters at hand are ruled out particularly within morphemes but not across morpheme boundaries. A reviewer notes that within morphemes, almost all English obstruent clusters are uniformly voiceless; some voiced or

---

24 I am grateful to reviewers for their insight into the issues discussed in this section.
25 See, e.g., the oft-cited case of exceptionality in Yine, formally known as Piro (Matteson 1965; Kisseberth 1970; Pater 2009), in which two homophonous affixes have different phonological effects.
26 As noted below, 4-year-old children have already acquired considerable knowledge of the special right-edge clusters that the English plural suffix permits—at least do[ŋz] and ca[tʃ], if not also w[ðz]. For the argument just made to go through, it must be the case that these children set up a new OO-faith constraint for every new sound-meaning pair; although they already know an English plural, this alien plural’s different phonological shape (‘-del’ vs. ‘-s’/-z/-əz’) must trigger the creation of its own OO-faith constraint.
mismatched clusters occur in Latinate words that would appear to be no longer polymorphemic (e.g. anecdote, abdicate), while compounds allow voicing to freely mismatch (blackboard, pigpen, etc.). Given that the experiment compared the same clusters in both morphological conditions, the L1 English grammar could have created the increased faithfulness to clusters in plurals vs. mass nouns if it already allowed more marked structure across a coda-onset cluster in morphologically complex, but not simple, words.

To build such a grammar, these 4-year-olds would need to have determined the morphological structure of all forms like blackboard, pigpen, bedbug, etc. If they have learned enough such words but not enough about their morphology to posit OO-faith explanations for their voicing, they will have demoted AGREEVOICE below IO-faith (akin to the superset grammar from Table 1), predicting that their performance should be equally good in both experimental conditions. It is true that if they have determined that AGREEVOICE must be obeyed only within morphemes, they might have an alternative reason to be faithful to the voicing in wutchdel’s cluster compared to that in zitchdin. Two caveats should be noted, however. First, these English voicing mismatches are overwhelmingly at the juncture between two full words (e.g. black and board), whereas the experimental clusters are clearly at a root+suffix boundary, and wug testing famously shows that young children are aware of the need to match voicing among obstruents in this position, as in dog[z] vs. cat[s], starting with Berko (1958.) Second, any asymmetries in the location of the error (coda vs. onset) are not explicable from L1 transfer, as no English word-medial voicing is repaired in onsets. Thus it is unlikely that L1 transfer alone could have produced the experimental effects.\footnote{Of course, the best way to resolve this issue would have been to carefully examine these children’s obstruent cluster production in English and somehow to further determine their grammaticality judgments on voicing-matched and mismatched clusters to compare with their performance in the artificial language. No systematic attempt to collect or analyze participants’ English production was made, in large part because the experiment was already long and sufficiently taxing for this age group, and the circumstances of the data collection did not permit multiple sessions with each participant. While the participants who made it into the study were necessarily talkative enough to engage in the task, the majority of their productions were in fact productions of the target items, but it is worth noting that the transcribers’ relistening to the participants’s English productions did not reveal any evidence of medial cluster reduction in English.}

A different version of this concern raised by a reviewer is the possibility that children interpreted one or the other of the morphological categories as in fact being compounds, possibly as already mentioned due to their heavy or superheavy initial syllables. If the plural suffix was interpreted as its own word, this would make the possibility of L1 transfer into a serious confound in explaining the data. As one attempt to determine the likelihood of this analysis, a naïve transcriber was asked to recode all the children’s token for stress, including secondary stress if any. The result of this impressionistic query was that all but two bisyllabic tokens were heard as stressed only on the first syllable; the two remaining tokens, one plural and one mass, were coded with stress on both syllables. This lack of secondary stress on the overwhelming majority of final syllables suggests that the compound interpretation was unlikely to be a source of the plural vs. mass noun differences.

With all of these considerations in hand, the next section returns to the study’s basic question—did participants’ production data show a bias for paradigm uniformity?—and interprets each child’s results within an OT grammar.
4.2. Stages of Medial Cluster Faithfulness

The data in Table 15 was used in section 3.3 to group participants into three groups, each of whose grammars will now be considered in turn. The first group can be labelled the Coda-Unfaithful group, as their error patterns in both plurals and mass nouns showed a clear tendency to limit unfaithfulness in medial clusters to the first syllable’s coda (e.g.: wutchdel → wut.del, zitchdin → zidʒ.din.) This pattern could be attributed to two different kinds of grammars. The first alternative is that despite having learned the suffix -del at some level, and being able to spontaneously apply it correctly in plural contexts, its phonological role in triggering a morphological need for paradigm uniformity was not sufficiently established in the course of the experiment. OO-del-faith constraints were not stabilized in some sense and so these learners remained in a premorphology state, which Table 4 showed to be one in which medial clusters are always reduced in coda rather than onset. An explanation of how the learner could both know and not know about the novel suffix -del in this way might lie with the mental resources required to implement an OO-faithful grammar: setting up a lexical entry for a closed class affix like ‘plural,’ constructing a morphologically complex input to the phonological grammar, and the like.

A somewhat more complicated alternative is that these learners’ grammars rank OO-faith above IO-faith-Onset, so that onsets are protected regardless of basehood or not, and that the markedness constraints at issue are still ranked above general faithfulness constraints, driving the coda repair. This option is illustrated in Tables 17 and 18.28

---

28While the issue of how learners ought to acquire intermediate stage rankings like this one is not addressed explicitly in this article (see discussion in section 5.2), this particular stage shown in Tables 17 and 18 is not easy to derive in any current error-driven learner of English, to my knowledge (cf. the stages predicted and exemplified in Jesney & Tessier 2011). Thus, I consider that the Coda-Unfaithful group is most likely to have been using a phonology that, with respect to OO-del, is premorphological awareness.
The second group seen in Table 15 was the Faithful group, in which medial clusters were most often not repaired. For these learners, general IO-faithfulness outranks both the markedness constraints that are violated by each cluster and any conflicting OO-faith constraints. (Section 5.2 will return to this final-state grammar and its properties).

It is the third group of participants, who can be called the Morphologically-Asymmetric group, who fuel the majority of the interpretation in this article. This group’s differential treatment of plurals compared to mass nouns could in principle demonstrate either of the post-morphology stages discussed in section 1.3, and the examination of these children’s productions below provides evidence of both.

Table 15 showed that these four children (E, A2, A3, and N) together produced 12/36 faithful initial syllable codas in plural nouns, but 25/27 faithful initial syllable codas in mass nouns. The four parts of Table 19 below illustrate the set of tokens for each Morphologically-Asymmetric child, on clusters that they produced in both morphological conditions. Each provides evidence that the first syllable codas of plurals were more faithful than those of mass nouns: the relevant clusters are underlined (numbers in brackets indicate the number of tokens) (see Table 19).

A good illustration of the morphological asymmetries in these data comes from A2’s treatment of the [b.d] cluster in Table 19b. In the mass noun gibdit, A2 reduces the coda [b] to a glottal stop—in Table 20 this is attributed to a markedness constraint against a sequence of two oral stops, and this repair is chosen over others, such as epenthesis (see Table 20).

On the other hand, A2’s tokens of this same cluster in the plural pobdel show two different treatments. Table 21 illustrates the alternative repair: epenthesis, which allows both input cluster segments to survive segmentally, yet both be syllabified as onsets. The same ranking from Table 20 can be used to create this alternative repair, if the \text{DEP} \gg \text{IDENT[PLACE]} \text{ ranking is subverted by the need to be OO-faithful to pobdel’s [b]}, via the base pob (see Table 21).

The other treatment of pobdel in A2’s data is full faithfulness, which suggests a small ranking variation from Table 21 to Table 22. In the latter, the markedness constraint *\text{ORALSTOP-ORALSTOP} has switched places with \text{DEP-IO}, ruling out the last remaining repair. As in the previous table, the presence of \text{IDENT-[PLACE]-OO} prevents the coda from debuccalizing as it did in the mass noun gibdit (see Table 22).

The fact that A2 produced the same plural word pobdel in (at least) two different ways over the course of a fifty-minute experiment is no surprise to any child language researcher, but this degree of variability (seen throughout the experiment) is not straightforward to capture in many frameworks. One simple approach that does minimal damage to the “classic” OT of Prince & Smolensky (1993/2004) is to assume that the grammar includes sets of constraints that are crucially unranked with respect to one another and whose total ordering is decided each time the grammar is used (see, e.g., Nagy & Reynolds 1995; Anttila 1997, 2002; Hayes 2000). Given this experiment’s somewhat large task demands, floating constraints may be an appealing way to model variability in the face of strained mental resources—on this latter point, see Davidson, Smolensky & Jusczyk (2004), Pater, Stager & Werker (2004), and Werker and Curtin (2005).}

\footnote{Assuming that glottal stop, without any constriction in the oral cavity, is not targeted by this constraint.}

\footnote{For a different take on how the grammar produces child variability, see Boersma & Levent (2000) and Curtin & Zuraw (2001); for a view of child variability created by the learning process rather than the grammar, see Becker & Tessier (2011).}
### TABLE 19
(a) E’s Comparable Clusters in Plural and Mass Nouns

<table>
<thead>
<tr>
<th>Mass</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Target</strong></td>
<td><strong>Child</strong></td>
</tr>
<tr>
<td>fad₃det</td>
<td>fad₃.det (4)</td>
</tr>
<tr>
<td>gamfdep</td>
<td>gam.det (1)</td>
</tr>
<tr>
<td></td>
<td>gamf.dep (2)</td>
</tr>
<tr>
<td></td>
<td>gam.det (1)</td>
</tr>
<tr>
<td>z1₃fd1₃n</td>
<td>z1₃zd1₃n (1)</td>
</tr>
</tbody>
</table>

(b) A2’s Comparable Clusters in Plural and Mass Nouns

<table>
<thead>
<tr>
<th>Mass</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Target</strong></td>
<td><strong>Child</strong></td>
</tr>
<tr>
<td>fad₃det</td>
<td>fat₃.dep (2)</td>
</tr>
<tr>
<td>gbt1det</td>
<td>gbt1.det (2)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c) A3’s Comparable Clusters in Plural and Mass Nouns

<table>
<thead>
<tr>
<th>Mass</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Target</strong></td>
<td><strong>Child</strong></td>
</tr>
<tr>
<td>gbt1det</td>
<td>gbt1.det (2)</td>
</tr>
<tr>
<td>gamfdep</td>
<td>gamf.dep (2)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>z1₃fd1₃n</td>
<td>z1₃zd1₃n (3)</td>
</tr>
<tr>
<td></td>
<td>z1₃zd1₃n (2)</td>
</tr>
</tbody>
</table>

(d) N1’s Comparable Clusters in Plural and Mass Nouns

<table>
<thead>
<tr>
<th>Mass</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Target</strong></td>
<td><strong>Child</strong></td>
</tr>
<tr>
<td>fad₃det</td>
<td>fa.tfa.det</td>
</tr>
<tr>
<td>z1₃fd1₃n</td>
<td>z1₃zd1₃n (2)</td>
</tr>
<tr>
<td></td>
<td>z1₃zd1₃n (2)</td>
</tr>
</tbody>
</table>
TABLE 20
A2’s Treatment of [b,d] in a Mass Noun: Coda Debuccalization

<table>
<thead>
<tr>
<th>/gtb_dt/</th>
<th>ID-ONS[PLACE]-IO</th>
<th>*ORALSTOP-ORALSTOP</th>
<th>DEP-IO</th>
<th>ID-[PLACE]-IO</th>
</tr>
</thead>
<tbody>
<tr>
<td>gtb_dt</td>
<td>!</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>gtb_dt</td>
<td>!</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>gr_ba_dt</td>
<td>!</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>~gti_dt</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 21
A2’s Treatment of [b,d] in a Plural: Epenthesis

<table>
<thead>
<tr>
<th>/pob_da_t/</th>
<th>ID-[PLACE]-OO</th>
<th>ID-[PLACE]-ONSET-IO</th>
<th>*ORALSTOP-ORALSTOP</th>
<th>DEP-IO</th>
<th>ID-[PLACE]-IO</th>
</tr>
</thead>
<tbody>
<tr>
<td>pob_da_t</td>
<td>!</td>
<td>!</td>
<td>!</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>po_da_t</td>
<td>!</td>
<td>!</td>
<td>!</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>pob_ta_t</td>
<td>!</td>
<td>!</td>
<td>!</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>~pob_ba_da_t</td>
<td></td>
<td>!</td>
<td>!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 22
A2’s Treatment of [b,d] in a Plural: Faithful

<table>
<thead>
<tr>
<th>/pob_da_t/</th>
<th>ID-[PLACE]-OO</th>
<th>ID-[PLACE]-ONSET-IO</th>
<th>DEP-IO</th>
<th>*ORALSTOP-ORALSTOP</th>
<th>ID-[PLACE]-IO</th>
</tr>
</thead>
<tbody>
<tr>
<td>~pob_da_t</td>
<td></td>
<td></td>
<td>*</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>po_da_t</td>
<td>!</td>
<td>!</td>
<td>*</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>po_ta_t</td>
<td>!</td>
<td>!</td>
<td>*</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>pob_ba_da_t</td>
<td></td>
<td>!</td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

5. IMPLICATIONS AND ALTERNATIVES

5.1. Alternative Accounts

Although sections 2 and 4 provided arguments against many reinterpretations or possible confounds in this study, there are at least two potential alternatives raised by reviewers that cannot be easily distinguished from its intended purpose. The first possibility is a perceptual difference between plural and mass nouns, from the participants’ perspective—perhaps mass nouns were less faithful in their first syllable codas because their first syllables had never been heard in isolation (being smaller than a single morpheme), so that their codas were not as well-learned as those in the same position of plural nouns, where the coda had indeed been heard word-finally as isolated in the singular.

It seems impossible to prevent this potential confound in the materials with the present methodology. One attempt to check the likelihood of this possibility post-hoc was the
retranscription of all the puppet (i.e., experimenter’s) ‘model productions’ of Bozdim’s words, by an additional, phonetically trained transcriber naïve to the purposes of the study. Her transcriptions revealed only three errors, in which the place of articulation of nasal and fricative in *gumfdep* and *naemfdel* was uncertain. The other piece of positive evidence available are the sample spectrograms given in the Appendix, which are used to illustrate the similar degree of ‘slight release’ of coda consonants in both bisyllabic conditions as claimed in section 2.4. In future work, it would be ideal to use a perceptual task to determine whether the participants were equally good at recognizing errors in the first syllable codas of both plural and mass nouns.

A different, representational interpretation raised by a reviewer is that base-final segments might be protected in the grammar not by OO-faith but instead by faithfulness to the edge of Prosodic Words, under the assumptions that English plural inflection sits outside the PWd, and that this representation is either transferred to this L2 (see especially Goad & White 2006) or else assumed in some way by default. This study was not designed to distinguish between these two kinds of faithfulness in the plural condition so their predictions cannot be teased apart here, but it should be noted that a PWord account will not explain all cases of attested innovative OO-faith in attested child phonologies, such as the preservation of base stress in early Hebrew mentioned in section 1.3.31

### 5.2. Moving from Initial to Final States

Over the course of this experiment, the language as presented to the learners contained words and clusters with no alternations or phonological repairs—that is, with no evidence of unfaithfulness to any marked properties. As discussed in section 4.2, this means that compared to the initial state, the target language is one in which general IO-faith constraints outrank any conflicting Markedness constraints:

(5) a. *Initial state, repeated from 3)*:

\[\text{OO-faithfulness} \gg \text{Markedness} \gg \text{IO-faithfulness}\]

b. *Necessary final state ranking*

\[\text{IO-faithfulness} \gg \text{Markedness}\]

While the ranking in (5b) produces the correct faithful outputs, it does not address the position of OO-faith or any more specific IO-faith constraints on voicing—in Table 23 these additional faithfulness constraints are shaded out to indicate their inability to affect the choice of winner (see Table 23).

As alluded to in section 1, it is ranking biases that will decide the position of these constraints in the final grammar. Since the initial state ranks OO-faith at the top of the hierarchy and the

31A somewhat related suggestion is that children parsed the –[dəl] suffix as a derivational affix or even a compounding root, by virtue of its exceptional English medial-cluster phonology. In other words, since the morpheme boundary of count plurals like *witchdel* do not agree for cluster voicing, and this puts them on a phonological par with English ‘-ship’ or ‘-dom’ and not plural [-z] or [-s], the child may have been led to assume a more derivational meaning for the suffix, perhaps ‘*witch-more, i.e., more than one witch* and therefore treated them as impervious to English voicing agreement. This interesting possibility cannot be ruled out, although it may not account for all of this study’s results—especially since the English frequency of some of these clusters is so low that it is hard to be sure that any 4-year-old knows they are possible in any English morphological context.
language provides no evidence to the contrary, a biased learner will leave OO-faith at the top of the hierarchy. Together, then, the evidence plus biases provides the end state in (6):

\[(6) \text{ Full final state ranking} \]
\[\text{OO-faithfulness} \gg \text{IO-faithfulness} \gg \text{Markedness} \]

Precisely how the learner gets from (5a) to (6) is a technical but certainly nontrivial matter. The predominant view is that learners use the errors they make in reproducing observed words to highlight the differences between the current and target grammars, and to use re-ranking algorithms that identify and capitalize on these errors to progress from initial to final states. An additional layer of complexity comes from the learner’s intermediate stages of acquisition and how they are derived in the learning process. Spelling out the mechanisms and algorithms that move from initial to intermediate to final stages is well beyond the scope of this article—but on the question of reaching the correct final state, in the non-stochastic version of OT adopted here, see especially Tesar & Smolensky (2000), Prince & Tesar (2004), and Hayes (2004); on the latter question of ensuring accurate intermediate stages, see especially Tessier (2009). Similar questions and answers have also been proposed within a stochastic view of constraint ranking (see especially Boersma & Hayes 2001; Pater 2008)—especially with respect to intermediate stages, whether in OT (e.g., Boersma & Levelt 2000, Curtin & Zuraw 2001, and Levelt & van der Vijver 2004) or more recently in Harmonic Grammar (Jesney & Tessier 2011).

6. CONCLUSIONS

This study’s results provide novel evidence that children rely on high-ranked OO-faith constraints over the course of phonological acquisition. When 4-year-old children were faced with difficult consonant clusters in a new language, their repairs to those clusters demonstrated preferences for OO-faithfulness over both Markedness and IO faithfulness.

Although many questions remain, the positive result is encouraging in at least two ways. First, it provides support for any learning theory that predicts the early emergence of OO-faithfulness, independent of data triggers from the target. A second encouraging result is the general indication that children are both willing and able to engage in artificial language learning of this type, particularly in learning new functional material like a plural suffix. This suggests a scarcely tapped data source for a range of experimental work on other aspects of learnability theory and its consequences. Careful experimental design and investigation of other experimental paradigms will provide better insight into how the methodology of artificial language learning can best be applied to the study of early phonological and morpho-phonological development.

<table>
<thead>
<tr>
<th>/wʌtʃ + daːt/ base [wʌtʃ]</th>
<th>ID-[VCE]-IO</th>
<th>AGREE-VOICE</th>
<th>ID-Onset [VCE]-IO</th>
<th>ID-[VCE]-OO</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 wʌtʃ,daːt</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wʌtʃ, daːt</td>
<td>*!</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wʌdʒ, daːt</td>
<td>*!</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

Thanks in largest part to Joe Pater, Heather Goad, and four anonymous reviewers for their innumerable improvements to this article—particularly one reviewer whose conscientious attention to detail prevented the final draft from including more than the usual number of glaring errors. Special thanks to the teachers, director, parents, and children at Cushman School, Amherst, MA, in 2003–2004 for their enthusiastic support and participation in the study, and further thanks to John McCarthy, Lyn Frazier, Shelley Velleman, John Kingston, Bruce Hayes, Kie Zuraw, Elan Dresher, Beth Hume, Michael Wagner, Michael Becker, Kathryn Flack, and audiences at MOT2005, HUMDRUM 2005, BUCLD30, and the University of Toronto. All remaining errors are the author’s alone.

REFERENCES


APPENDIX: Sample Spectrograms (see Section 5.1)

FIGURE 1  Sample productions of [tʃ,d] in plural (top) and mass (bottom) items (color figure available online).

FIGURE 2  Sample productions of [mʃ,d] in plural (top) and mass (bottom) items (color figure available online).