JOURNAL ARTICLE

Inductive learning of locality relations in segmental phonology

Kevin McMullin[1] and Gunnar Ólafur Hansson[2]

[1] Department of Linguistics, University of Ottawa, Ottawa, ON, CA

[2] Department of Linguistics, University of British Columbia, Vancouver, BC, CA

Corresponding author: Kevin McMullin (kevin.mcmullin@uottawa.ca)

This paper reports on a series of artificial grammar learning experiments focused on locality relations in patterns of long-distance consonant agreement (harmony) and disagreement (dissimilation). Participants in experimental conditions were exposed to dependencies affecting stem-suffix pairs of liquids at either a short-range (transvocalic, CVCV<u>L</u>V-<u>L</u>V) or medium-range (beyond-transvocalic, CV<u>L</u>VCV-<u>L</u>V) distance. Two experiments used a poverty of stimulus paradigm, offering no information about the other distance level; participants interpreted short-range interaction as a strictly transvocalic dependency but medium-range interaction as unbounded, generalizing to other distances. Two experiments employed a 'rich stimulus' paradigm, where training data unambiguously indicated the absence of any dependency at the other distance; this enabled probing of specific locality patterns, in particular strictly beyond-transvocalic dependencies. The constraint-based Agreement by Correspondence model of non-adjacent consonant interactions predicts such patterns to be possible for dissimilation but not harmony. The results do not support this hypothesis: Participants seem to have serious difficulty learning strictly beyond-transvocalic dependencies of either kind. Our findings are more consistent with recent proposals that the space of learnable phonotactic restrictions is delimited by the Tier-based Strictly 2-Local class of formal languages. Strictly transvocalic and unbounded dependencies lie within this region, whereas strictly beyond-transvocalic dependencies are more complex, falling outside the learner's hypothesis space.

Keywords: learning bias; artificial grammar learning; locality relations; harmony; dissimilation; complexity; learnability

1. Introduction

Not all logically possible sound patterns are attested in the world's languages, and there is a striking amount of uniformity among those that are. An important goal for phonology is therefore to explain the cross-linguistic typology of sound patterns. The properties of individual languages, and thereby the overall typology, are shaped over time by various forces that can affect the intergenerational transmission of phonological systems (Blevins, 2004; Hansson, 2008; Kiparsky, 2008; Moreton, 2008; Ohala, 1983). On the one hand, the motor-planning, articulatory, aerodynamic, acoustic, and auditory-perceptual demands of speech production and perception skew the variability in the speech signal, and the listener's parsing of that signal, in consistent and largely language-independent ways. As systematic sound patterns emerge from this pool of variation through phonologization (Hyman, 1976; Kiparsky, 2015; Yu, 2013), such *channel biases* (Moreton, 2008) explain the predominance of synchronic phonological patterns that are phonetically 'natural' (Hansson, 2008).

Equally important are *inductive biases* that govern the learning of linguistic patterns: limitations on the hypothesis space available to the human learner, and heuristics by which

the learner traverses that space. Indeed, without inductive biases learning is impossible (Mitchell, 1990; Pinker, 1979). Such biases can in principle be domain-specific or domain-general (see e.g., Moreton, 2012) and either categorical ('hard') or probabilistic ('soft'). For example, the Universal Grammar (UG) notion familiar from generative models amounts to a set of domain-specific and (largely) categorical inductive biases, delimiting the space of possible (i.e., learnable) linguistic systems. In recent years, a growing body of research has used *artificial grammar learning* experiments (Reber, 1967, 1989) to probe for inductive biases in a controlled laboratory setting (e.g., Baer-Henney, Kügler, & van de Vijver, 2015; Baer-Henney & van de Vijver, 2012; Carpenter, 2010; Cristia, Mielke, Daland, & Peperkamp, 2013; Finley, 2011, 2012, 2015; Finley & Badecker, 2009; Gallagher, 2013; Kapatsinski, 2010; Kuo, 2009; Lai, 2015; Linzen & Gallagher, 2017; Moreton, 2008, 2012; Peperkamp & Dupoux, 2007; Pycha, Nowak, Shin, & Shosted, 2003; White, 2014; Wilson, 2003, 2006; for overviews, see Culbertson, 2012; Moreton & Pater, 2012a, 2012b). In the literature on artificial phonology learning, two main types of inductive biases have been proposed, which appear to influence the ease with which a learner is able to acquire sound patterns. First, *formal complexity* biases (Moreton & Pater, 2012a) relate to structural properties of the pattern itself; sound patterns that are simpler (by some independent criteria) are learned better than more complex ones. The second category, *substantive* biases (Moreton & Pater, 2012b), favour phonetically motivated (natural) patterns over phonetically arbitrary (unnatural) ones. In terms of their impact on phonological typology, the effects of substantive biases thus overlap those arising from channel bias. The existence and relative importance of substantive biases continues to be a topic of debate (Beguš, 2018; Glewwe, 2019; Glewwe et al., 2018; Greenwood, 2016; Hayes & White, 2013; Moreton & Pater, 2012b; Myers & Padgett, 2014).

The purpose of this paper is twofold. First, we report on a series of four artificial grammar learning experiments that investigate the role of inductive bias in the learning of long-distance consonant agreement (harmony) and disagreement (dissimilation). Second, we assess theoretical treatments of locality relations in such sound patterns in light of our results as well as the cross-linguistic typology. We argue that the Agreement by Correspondence model of long-distance segmental interactions (Bennett, 2015b; Hansson, 2010; Rose & Walker, 2004) makes predictions that are not supported by our findings, neither with respect to the learnability of specific patterns nor the inferences that learners draw from ambiguous evidence. These questionable predictions especially relate to differences between long-distance assimilation and dissimilation. We also consider various formal language theoretic characterizations of long-distance phonotactics and argue that the Tier-based Strictly 2-Local class (TSL$_2$; Heinz, Rawal, & Tanner, 2011) offers a straightforward account of the typology and our experimental findings; alternative formal language classes within the subregular hierarchy (Jäger & Rogers, 2012; McNaughton & Papert, 1971; Rogers & Pullum, 2011) are either too simple or too complex to account for the observed data.

1.1. Locality relations in long-distance phonotactics

The cross-linguistic typology of consonant harmony (Hansson, 2010; Rose & Walker, 2004) reveals a dichotomy with respect to locality relations. Attested harmony patterns are either strictly *transvocalic*, affecting only those consonant pairs that fall within an adjacent-syllable window (…CV.C…), or else *unbounded*, holding at any distance within the relevant domain, even when other consonants intervene. The sibilant harmony found in several languages of the Omotic family (a branch of Afro-Asiatic), spoken in southern Ethiopia, illustrates the distinction. Transvocalic locality is exemplified by Koorete (a.k.a. Koyra; North Omotic). As shown in (1), the causative suffix /-us/, which surfaces faithfully in (1a), harmonizes with a preceding [–anterior] sibilant when a single vowel intervenes, as in (1b). However,

harmony is not enforced across greater distances (i.e., when another consonant intervenes), as seen in (1c).

(1)   Transvocalic sibilant harmony in Koorete (Koyra; Hayward, 1982)
    a.   Causative suffix /-us/
        /pug-us-/   pugus-   'cause to blow'
        /kes-us-/   kesus-   'cause to go out'

    b.   Transvocalic contexts: Harmony
        /tutʃ'-us-/   tutʃ'uʃ-   'cause to shut'
        /paʃ-us-/   paʃuʃ-   'cause to cover up'

    c.   Beyond-transvocalic contexts: No harmony
        /ʃoh-us-/   ʃohus-   'wash (tr.)'
        /tʃ'aːn-us-/ tʃ'aːnus- 'cause to load'

In the related language Aari (South Omotic), by contrast, this sibilant harmony holds in beyond-transvocalic contexts as well. This is seen in the examples in (2), which involve the perfective suffix /-s/ (cf. /giʔ-s-it/ → [giʔ.sit] 'I hit,' /giʔ-er-s-it/ → [gi.ʔer.sit] 'I was hit'; Hayward, 1988, p. 290).

(2)   Unbounded sibilant harmony in Aari (Hayward, 1990)
    a.   /tʃ'aːq-s-it/   tʃ'aːq.ʃit   'I swore'
    b.   /ʃed-er-s-it/   ʃe.der.ʃit   'I was seen'
    c.   /ʒaːg-er-s-e/   ʒaː.ger.ʃe   'it was sewn'

There are two competing interpretations of the locality criterion that defines transvocalic harmony. A common characterization is in terms of *syllable adjacency*: Harmony applies when the consonants are tautosyllabic or in adjacent syllables, but not further apart (Bennett, 2015a, 2015b; Odden, 1994; Piggott, 1996; Rose & Walker, 2004; Walker, 2000; cf. also Suzuki, 1998; Wayment, 2009). An alternative criterion, implicit in the term 'transvocalic' used here, is *consonant-tier adjacency*: Harmony applies if and only if no other consonant intervenes (Hansson, 2010; Jurgec, 2011; McMullin, 2016; McMullin & Hansson, 2015). Clear-cut evidence bearing on the choice between these interpretations is hard to come by (though consonant-tier adjacency seems slightly better supported; Hansson, 2010, pp. 176–178). We will largely side-step this issue in what follows. For example, our stimuli all consist of strings of CV syllables; all consonant-tier-adjacent C…C sequences are thus also syllable-adjacent and vice versa.

1.2. Explaining the typology

The transvocalic versus unbounded dichotomy amounts to an implicational universal for long-distance assimilation. As the schematic overview in **Table 1** shows, interaction at beyond-transvocalic distances entails interaction in transvocalic contexts, but not vice

Table 1: Typology of locality relations in consonant harmony (Hansson, 2010; Rose & Walker, 2004).

Locality pattern	Distance			Attested?
	$C_x V.C_y V$	$C_x V.CV.C_y V$	$C_x V.CV.CV.C_y V$	
No harmony	−	−	−	yes
Transvocalic harmony	+	−	−	yes
Unbounded harmony	+	+	+	yes
Beyond-transvocalic harmony	−	+	+	no

versa. Thus, consonant harmony never follows a *strictly beyond-transvocalic* locality pattern, whereby consonants in …CV.C… contexts are allowed to disagree in the feature in question while assimilation is enforced for more distant consonant pairs.

Explanations for this typological gap can potentially be sought in two distinct domains. The absence of strictly beyond-transvocalic harmony could be due to such patterns being *diachronically* inaccessible by the processes through which non-adjacent assimilatory interactions develop; we consider this hypothesis in Section 1.2.1. Alternatively (or in addition), such sound patterns may be *synchronically* disfavoured or impossible. That is, with respect to the hypothesis space or heuristics available to learners, such patterns are either entirely inaccessible or vulnerable to reinterpretation (imperfect learning) and thus inherently unstable. The hypothesized *learning bias* in question could be either categorical or probabilistic, and either domain-specific (perhaps to phonological patterns specifically; see Heinz & Idsardi, 2011; Lai, 2012) or attributable to some domain-general aspect of human pattern recognition. In Sections 1.2.2 and 1.2.3 we discuss two different approaches to synchronic explanations along these lines.

1.2.1. Diachronically inaccessible?

Whether non-adjacent consonant assimilation patterns emerge from the misperception of coarticulatory effects (cf. Blevins, 2004; Ohala, 1993, 1994) or from motor-planning errors (Garrett & Johnson, 2013), it is reasonable to suppose that such sound changes arise in relatively short-range …CV.C… contexts. If this is where every non-adjacent dependency ultimately originates, then all attested cases of unbounded dependencies must by definition be secondary, resulting from extension (overgeneralization) of an originally transvocalic pattern to longer-range contexts (Dolbey & Hansson, 1999). This would be a straightforward instance of analogical change through imperfect learning (Kiparsky, 1968): A learner, encountering evidence that the configuration …$C_a VC_b$… is disallowed, wrongly concludes that C_a…C_b sequences are banned in general, rather than just when separated by a vowel. (Indeed, we find suggestive evidence of such generalization in our experiments; see Section 4.4 below).

This conjectural account of the diachronic relationships among locality types receives support from cases where the transvocalic versus unbounded dichotomy is instantiated by cognate sound patterns in closely related languages. For example, Bantu nasal consonant harmony is transvocalic in the vast majority of cases (e.g., Lamba, Ndonga, Bemba, Luba, Herero) but unbounded in a small cluster of languages (Yaka, Kongo, Kimbundu, and to some extent Suku; see Hansson, 2010, pp. 85–92). As the latter constitute an areally and perhaps genetically coherent group within Bantu (so-called Zone H), the unbounded variant must be a later extension of an originally transvocalic pattern. The same may well apply to the historical relationship among the Omotic sibilant harmony systems illustrated in (1) and (2) above.

By contrast, there is no obvious mechanism by which harmony would systematically cease to hold in transvocalic contexts while continuing to be enforced at longer distances; this would be required in order for a strictly beyond-transvocalic pattern to develop out of a previously unbounded one. Consider, for example, that the breakdown of harmony (in a given context) might come about by an influx of loanwords that violate the harmony restriction (in that context). In order for this to be a possible mechanism for transitioning from an unbounded to a strictly beyond-transvocalic harmony, we would have to imagine a scenario where the corpus of loanwords in question does contain a significant number of items with disharmonic …CVC… co-occurrences but, by sheer happenstance, contains no (or strikingly few) instances of such co-occurrences involving longer-distance …C…C… pairs. A shift directly from strictly transvocalic harmony to strictly beyond-transvocalic

harmony would involve an even more counterintuitive leap: losing harmony in all contexts where it previously applied, while simultaneously innovating harmony in those contexts where it did not.

By the diachronic hypothesis, strictly beyond-transvocalic harmony is not so much synchronically impossible as diachronically inaccessible: The necessary conditions for it to emerge would involve a complex and fortuitous series of events, rendering its typological probability of attestation effectively zero (for a similar argument regarding certain 'unnatural' patterns, cf. Beguš, 2018). If this diachronic inaccessibility were the *sole* explanation for the typological gap, we would have every reason to expect human learners to be fully capable of learning strictly beyond-transvocalic harmony, were they exposed to it.

1.2.2. Universal Grammar?

In generative treatments of non-adjacent assimilation, the transvocalic versus unbounded dichotomy, and hence the implicational relationship between transvocalic and strictly beyond-transvocalic contexts, has typically been built into the formal machinery (representations, parameters, constraints) of phonological grammars. As a particularly well-articulated example, we will consider the constraint-based Agreement by Correspondence (ABC) model of long-distance consonant assimilation (Bennett, 2015b; Hansson, 2010; Rose & Walker, 2004) within Optimality Theory (Prince & Smolensky, 2004). For a rule-based autosegmental approach with similar characteristics, see Odden (1994).

In the ABC model, similarity (in terms of shared features) can cause co-occurring consonants to enter into an abstract structural relation of *correspondence* (indicated by subscripting, e.g., $[s_x a \int_x a]$ with correspondence versus non-corresponding $[s_x a \int_y a]$). Various demands can be placed on segments linked by correspondence; these include requiring that they *agree* in some feature [F] (e.g., favouring $[\int_x a \int_x a]$ or $[s_x a s_x a]$ over disharmonic $[s_x a \int_x a]$). If both the demand for correspondence between similar segments and the demand for agreement in [F] among correspondents are ranked higher than faithfulness to input [±F] specifications, harmony in [F] will result.

In the simple case, this mechanism yields unbounded harmony, since none of the constraints mentioned refer to the distance between the (potentially) corresponding consonants. To derive strictly transvocalic harmony within the ABC model, two alternative mechanisms have been adopted. Building on earlier proposals for a locality parameter on autosegmental rules (Odden, 1994; Piggott, 1996), Rose and Walker (2004) propose a constraint PROXIMITY (further refined as CC-SYLLADJ in Bennett, 2015b), which requires correspondent segments to be no further apart than in adjacent syllables. When high-ranked, this constraint precludes the establishment of a correspondence relation for beyond-transvocalic …C…C… pairs; consequently, agreement will not be enforced in such cases, only for transvocalic/syllable-adjacent …CV.C… pairs. A different approach is to include constraints that explicitly call for correspondence (for the relevant segment types) in transvocalic contexts (Hansson, 2010; see also Bennett, 2015b). For example, a constraint like CORR[+strid]$_{CVC}$ would require correspondence for transvocalic sibilant pairs ($[\int Vs]$, $[sV\int]$, $[sVs]$, $[\int V\int]$), while making no such demand of beyond-transvocalic instances of $[\int…s]$, $[s…\int]$, etc. We return to the subtle differences between these two formal mechanisms in Section 1.3 below.

In the ABC model, strictly *beyond*-transvocalic harmony cannot be generated under any ranking of the posited constraints; it falls outside the factorial typology defined by the model. This amounts to a 'hard' or categorical (and domain-specific) learning bias. Learning a sound pattern is to learn the appropriate ranking (or perhaps weighting; Pater, 2009) of the constraints provided in UG. Since no ranking will capture the pattern, it is by definition not learnable. Learners exposed to evidence of a strictly beyond-transvocalic harmony are expected to fail; either they will misinterpret the dependency pattern as

unbounded (generalizing from beyond-transvocalic to transvocalic contexts) or they will fail to learn any such pattern at all (effectively treating it as noise in the data).

1.2.3. Language-theoretic complexity?

A different approach is to view the transvocalic versus unbounded dichotomy, and the absence of strictly beyond-transvocalic harmony, from the perspective of computational complexity and formal language theory. Heinz (2010) shows how, within the subregular hierarchy of formal languages (Jäger & Rogers, 2012; McNaughton & Papert, 1971; Rogers & Pullum, 2011), transvocalic and unbounded consonant harmony fall within the classes of Strictly k-Local (SL_k) and Strictly 2-Piecewise (SP_2) languages, respectively. Each of these can be acquired by simple learning algorithms with desirable computational properties, tracking n-grams (of length k) in the SL_k case and precedence relations (between pairs of segments, regardless of distance) in the SP_2 case. A slightly more expressive class, the Tier-based Strictly 2-Local (TSL_2) languages (Heinz et al., 2011), encompasses not only transvocalic and unbounded patterns but also ones that involve blocking by certain intervening segments (McMullin, 2016). The TSL_2 class has favourable learnability properties as well (Jardine & Heinz, 2016; Jardine & McMullin, 2017).

As a definition of a phonotactic dependency, a TSL_2 formal language (stringset)—or, rather, the grammar generating such a language—consists of a set R of prohibited bigrams (adjacent pairs) of segments (e.g., $R = \{$*sʃ, *ʃs$\}$), together with a set of segments T that defines the *tier* on which adjacency is assessed and bigrams tracked (e.g., $T = \{$s, ʃ$\}$). The difference between transvocalic and unbounded harmony can be derived by varying the contents of T (McMullin, 2016). The example just given, with $R = \{$*sʃ, *ʃs$\}$ and $T = \{$s, ʃ$\}$, defines an unbounded harmony: Disharmony is banned in contexts like *[saʃa] and *[samaʃa] alike, because [m] is not in T and the [s…ʃ] sequence is thus tier-adjacent (constituting a bigram [sʃ] on the tier). However, if T is defined as including not just those consonants that figure in R but all consonants, $T = \{$s, ʃ, m, k, …$\}$, a form like [samaʃa] no longer contains an offending bigram [sʃ] on the tier, only the permissible bigrams [sm] and [mʃ]. The result is a strictly transvocalic harmony, applying in contexts like *[saʃa] but not [samaʃa], [samakaʃa], etc.

By contrast, a strictly *beyond*-transvocalic dependency, such as where disharmony is tolerated in [saʃa] but not in *[samaʃa], *[sakamaʃa], etc., would not be TSL_k for any k (nor would it be SL_k or SP_k). To distinguish [saʃa] from *[saCaʃa], where C is any non-sibilant, the tier T must include all consonants ($T = \{$s, ʃ, m, k, …$\}$). It is then necessary for R to include *[smʃ], *[skʃ], *[skmʃ], etc. (but crucially not [sʃ], which is permitted). However, since the grammar must be able to detect a disharmonic [s…ʃ] pair across a potentially unlimited number of intervening consonants (so as to rule the word out as ungrammatical), no upper bound on k will suffice. No matter what k is set to, a word of the form *…sV(CV)$^{k-1}$ʃ… will always be erroneously classified as well-formed. This is because the shortest n-gram (k-factor) of such a word on the consonant tier that contains both members of the long-distance pair [s…ʃ] is [sC$_1$C$_2$…C$_{k-1}$ʃ], the length of which is $k + 1$. By definition, the set of prohibited k-factors (R) cannot include strings of length greater than k.

Rather, a strictly beyond-transvocalic dependency would belong to (tier-based instantiations of) the superordinate class of Locally Testable (LT) languages or the even more encompassing Locally Threshold-Testable (LTT) class (Jäger & Rogers, 2012; Rogers & Pullum, 2011). The LT_k class is defined in terms of Boolean operations over sets of n-grams (of length k); LTT characterizations may also *count* the occurrences of each n-gram (up to some threshold). In the hypothetical strictly beyond-transvocalic sibilant harmony scenario above, an illegal word is one that simultaneously contains a member of {Cs, #s} and a member of {Cʃ, #ʃ} among its bigrams on the tier of all consonants (and word boundaries; C is shorthand for any non-sibilant). This is the case for *[samaʃa] and *[sakamaʃa], each of which contains both [#s] and [mʃ] on the tier, but not for [saʃa]

(where the only bigrams on the tier are [#s], [sʃ], and [ʃ#]). Evidence suggests that the LT and LTT regions of the subregular hierarchy may be inaccessible to human phonotactic learners. For example, Lai (2015) found that English adults failed to learn a 'first-last' sibilant harmony, where words with the structure #s…ʃ# or #ʃ…s# are banned, but where sibilants are otherwise unrestricted (such that, crucially, both #s…ʃ…s# and #s…s…s# are permitted). Such a pattern is also Locally Testable, as a word may not simultaneously include the bigrams [#s] and [ʃ#] on the tier of sibilants (or that of all consonants). Participants tended instead to infer an unbounded harmony, ruling out all words with disagreeing sibilants, regardless of whether these constituted a first-last pair.

From this perspective, too, the absence of strictly beyond-transvocalic harmony from the cross-linguistic typology is due to a learning bias that constrains the formal properties of synchronic sound patterns, and hence of phonological systems. However, it is less clear in this case whether that bias is categorical or merely probabilistic. On the latter interpretation, non-TSL$_2$ (i.e., LT or LTT) patterns are merely harder to learn than TSL$_2$ ones, perhaps requiring exposure to a greater amount, or different kinds, of evidence (for an example of 'soft' biases in phonological learning, see White, 2017). It is also less clear whether this complexity bias is necessarily a domain-specific property of linguistic/phonological pattern learning or whether it might be attributable to domain-general aspects of human cognition.

Traditionally, formal-language-theoretic characterizations of phonotactic dependencies do not make direct reference to phonological *features* when characterizing sets of segments or prohibited bigrams. However, we stress that this does not bear on any of the predictions relevant for this paper. We could, for example, simply define the space of possible tiers and co-occurrence restrictions using featural descriptions (natural classes), and assume that this space is provided to the learner a priori, in much the same way that the constraints of the ABC model are assumed to be given as part of UG. Doing so would further reduce the learner's hypothesis space (i.e., to the types of feature-based patterns that seem to be favoured cross-linguistically). Crucially, this would not affect the range of locality relations predicted to be possible within the TSL$_2$ region (or any other class of formal languages).

1.3. *Harmony and dissimilation: Diverging predictions*

The preceding discussion has revolved entirely around non-adjacent consonant *assimilation* (consonant harmony). One question raised in recent research is to what extent the same generalizations hold for non-adjacent *dissimilation*. For various segment types, both kinds of interaction are attested. For example, liquid harmony is found in Bukusu (Bantu; Hansson, 2010; Odden, 1994), where one of its manifestations is the applicative suffix /-il/ surfacing as [-ir] when the preceding stem contains [r] (e.g., /-rum-il-a/ → [-rumira] 'send for,' cf. /-xam-il-a/ → [-xamila] 'milk for'). Liquid dissimilation is robustly attested as well, for instance in Georgian (Kartvelian; Fallon, 1993; Odden, 1994), where the ethnonym suffix /-uri/ surfaces as [-uli] when the preceding stem contains [r] (e.g., /aprik'-uri/ → [aprik'uli] 'African,' cf. /dan-uri/ → [danuri] 'Danish'). In the Bukusu case, [r…l] sequences are avoided in favour of [r…r], while the exact opposite is true for Georgian.

With respect to the characterization of long-distance phonotactic dependencies as TSL$_2$ stringsets, harmony and dissimilation are equivalent as regards locality relations. A TSL$_2$ grammar specifying $R = \{$*rr, *ll$\}$ and $T = \{l, r\}$ amounts to an *unbounded* dissimilation pattern, banning [r…r] and [l…l] co-occurrences at any distance, regardless of what intervenes. If the tier is defined as including not just liquids but all consonants, $T = \{l, r, p, t, k, …\}$, a *transvocalic* dissimilation pattern is generated: [rVr] and [lVl] sequences are banned, whereas [r…r] and [l…l] are permitted as long as at least one consonant intervenes (e.g., grammatical [ri.pa.ru] versus ungrammatical *[pa.ri.ru]). To translate these TSL$_2$ grammars into ones generating unbounded and transvocalic liquid harmony,

respectively, one need only define the set of prohibited bigrams as $R = \{*rl, *lr\}$ (cf. the sibilant harmony example described earlier).

In the ABC model, however, where dissimilation is interpreted as 'correspondence avoidance' (Bennett, 2015b), the equivalence between harmony and dissimilation breaks down. As harmony (agreement in [F] between correspondents) presupposes a correspondence relation between the two consonants, any restrictions on its application are expected to confine harmony to contexts where correspondence is favoured. Dissimilation, on the other hand, emerges in ABC as a means of evading the need for establishing a correspondence relation in the first place; reducing the similarity of the two consonants renders them free to *not* stand in correspondence. Dissimilation is therefore expected to gravitate toward contexts where correspondence is *disfavoured* (or where it is not explicitly required). One corollary of this viewpoint is that various mismatches are predicted to hold between the typology of harmony and that of dissimilation. Bennett (2015b) provides cross-linguistic support for several such mismatch predictions, particularly relating to structural factors (e.g., harmony tending to occur within domains, dissimilation across domain edges).

One such asymmetry between harmony and dissimilation is expected to arise with respect to locality relations. For consonant harmony, strictly transvocalic locality can arise in ABC by either of two mechanisms, as noted in Section 1.2.2. First, a high-ranked constraint demanding correspondence regardless of distance (for the relevant consonant types) may be overridden by even higher-ranked PROXIMITY/CC-SYLLADJ, which explicitly bans correspondence except for …CV.C… consonant pairs. Alternatively, the high-ranked constraint that demands correspondence may itself be one that singles out …CVC… pairs specifically. Either way, the result will be the same: The only consonant pairs subject to harmony (i.e., featural agreement under correspondence) are transvocalic ones. In the case of dissimilatory interactions, these same formal mechanisms that give rise to transvocalic harmony can instead have the effect of limiting dissimilation to consonant pairs *beyond* the correspondence-favouring …CV.C… window. In this way, the ABC model can generate a strictly beyond-transvocalic locality pattern for dissimilation.

In (3)–(4) we show how this results from the correspondence-delimiting constraint PROXIMITY/CC-SYLLADJ. The constraint CORR-LIQUID[αlateral] demands correspondence for pairs of co-occuring identical liquids ([r…r], [l…l]), regardless of distance, while lower-ranked CORR-LIQUID does the same for all liquid pairs (also [r…l], [l…r]); co-indexation denotes correspondence. IDENT-IO[±lateral] penalizes changes to [±lateral] input specifications.

(3)　　Transvocalic contexts with PROXIMITY: No dissimilation (/rVr/ → [rVr])

	/pariru/	PROXIMITY	CORR-LIQ[αlat]	IDENT-IO[±lat]	CORR-LIQ
a.	pa.r$_x$i.r$_y$u		*!		*
b. ☞	pa.r$_x$i.r$_x$u				
c.	pa.r$_x$i.l$_y$u			*!	*

(4)　　Beyond-transvocalic contexts with PROXIMITY: Dissimilation (/r…r/ → [r…l])

	/riparu/	PROXIMITY	CORR-LIQ[αlat]	IDENT-IO[±lat]	CORR-LIQ
a.	r$_x$i.pa.r$_y$u		*!		*
b.	r$_x$i.pa.r$_x$u	*!			
c. ☞	r$_x$i.pa.l$_y$u			*	*

Furthermore, a transvocalic pair of liquids that already disagrees for [±lateral] will surface faithfully, as shown in (5); the winning candidate does not violate any of the four constraints. The ranking shown here therefore results in a language where the co-occurrence of liquids is *unrestricted* in transvocalic contexts, while liquid dissimilation is enforced at beyond-transvocalic distances.

(5)   Transvocalic contexts with PROXIMITY: No interaction (/rVl/ → [rVl])

	/parilu/	PROXIMITY	CORR-LIQ[αlat]	IDENT-IO[±lat]	CORR-LIQ
a.	pa.r$_x$i.l$_y$u				*
b. ☞	pa.r$_x$i.l$_x$u				
c.	pa.r$_x$i.r$_x$u			*!	

Strictly beyond-transvocalic dissimilation can likewise result from the other formal mechanism that generates strictly transvocalic harmony: Special locality-restricted versions of CORR constraints, such as CORR-LIQ$_{\text{CVC}}$ or CORR-LIQ[αlat]$_{\text{CVC}}$ alongside their unrestricted counterparts shown in the above tableaux. To see how this is true, we must first introduce another, independently motivated aspect of the ABC constraint repertoire. In order to derive truly *unbounded* dissimilation in ABC (e.g., affecting [r…r] and [l…l] regardless of distance or other contextual factors), some constraint or constraints must exist whose effect is to ban correspondence in *all* contexts. For example, this can result from the combined forces of CC-EDGE(σ), which requires correspondents to be tautosyllabic, and CC-SROLE, which requires correspondents to occupy matching syllable constituents (onset…onset, coda…coda); for justification for both of these, see Bennett (2015b). Since these demands are mutually contradictory (at least for non-adjacent consonants), the net result of enforcing both of these constraints is that no two consonants can ever stand in correspondence. In the tableaux below, we will use the shorthand label CC-NOCORR to represent such an across-the-board ban on correspondence.

With these ingredients in place, the tableaux in (6)–(7) show how, given the right constraint ranking, liquid dissimilation will be enforced in beyond-transvocalic contexts while leaving transvocalic liquid-liquid pairs unaffected. Tableau (8) further shows that transvocalic pairs are unrestricted; they surface faithfully regardless of whether they agree (6) or disagree (8) in the input.[1]

(6)   Transvocalic contexts with CORR-LIQ$_{\text{CVC}}$: No dissimilation (/rVr/ → [rVr])

	/pariru/	CORR-LIQ$_{\text{CVC}}$	CC-NOCORR	CORR-LIQ[αlat]	IDENT-IO[±lat]	CORR-LIQ
a.	pa.r$_x$i.r$_y$u	*!		*		*
b. ☞	pa.r$_x$i.r$_x$u		*			
c.	pa.r$_x$i.l$_y$u	*!			*	*

---

[1] The outcome in (8) further depends on a ranking IDENT-IO[±lat] ≫ CC-IDENT[±lat]; the latter constraint demands that correspondents agree in [±lateral], and thus favours candidate (d) [pa.r$_x$i.r$_x$u] over (b) [pa.r$_x$i.l$_x$u]. Reversing this ranking produces a language in which beyond-transvocalic liquid dissimilation co-exists with transvocalic liquid harmony. Full details of the range of locality patterns generated by the inventory of ABC constraints considered here are provided in Section 9.2.1 and in the Appendix (on-line supplement).

(7)   Beyond-transvocalic contexts with CORR-LIQ$_{CVC}$: Dissimilation (/r…r/ → [r…l])

	/riparu/	CORR-LIQ$_{CVC}$	CC-NOCORR	CORR-LIQ[αlat]	IDENT-IO[$\pm$lat]	CORR-LIQ
a.	r_xi.pa.r_yu			*!		*
b.	r_xi.pa.r_xu		*!			
c. ☞	r_xi.pa.l_yu				*	*

(8)   Transvocalic contexts with CORR-LIQ$_{CVC}$: No interaction (/rVl/ → [rVl])

	/parilu/	CORR-LIQ$_{CVC}$	CC-NOCORR	CORR-LIQ[αlat]	IDENT-IO[$\pm$lat]	CORR-LIQ
a.	pa.r_xi.l_yu	*!				*
b. ☞	pa.r_xi.l_xu		*			
c.	pa.r_xi.r_yu	*!		*	*	*
d.	pa.r_xi.r_xu		*		*!	

In sum, we have a set of diverging predictions as regards the possibility, and hence learnability, of different locality patterns in non-adjacent consonant assimilation and dissimilation, summarized in **Table 2**. It should be noted that the robustly attested pattern of strictly *transvocalic* dissimilation can only be generated in ABC if the constraint set includes locality-restricted CORR constraints such as CORR-LIQ[αlat]$_{CVC}$ or CORR-LIQ$_{CVC}$; Bennett (2015b) argues for the inclusion of such constraints (alongside the PROXIMITY/CC-SYLLADJ constraint) for precisely this reason. It is important to note, however, that regardless of whether the ABC constraint repertoire includes both of these constraint types or just one of the two, strictly *beyond-transvocalic* dissimilation is generated as part of the factorial typology, and is hence predicted to be part of the learner's hypothesis space.

    From the point of view of language-theoretic complexity, if long-distance phonotactic dependencies are indeed confined to the TSL$_2$ region, we expect strictly beyond-transvocalic locality to be impossible for assimilation and dissimilation alike. We note that the same prediction holds true for non-correspondence-based generative analyses of dissimilation in terms of segment co-occurrence constraints (such as the Obligatory Contour Principle,

Table 2: Predictions of possible (✓) versus impossible (✗) locality patterns in non-adjacent assimilation/harmony and dissimilation. TSL$_2$ = Tier-based Strictly 2-Local; L(T)T = Locally (Threshold) Testable; ABC = Agreement by Correspondence.

Dependency	Locality pattern	Hypotheses		
		TSL$_2$	L(T)T	ABC
Assimilation	Unbounded	✓	✓	✓
	Strictly transvocalic	✓	✓	✓
	Strictly beyond-transvocalic	✗	✓	✗
Dissimilation	Unbounded	✓	✓	✓
	Strictly transvocalic	✓	✓	✓
	Strictly beyond-transvocalic	✗	✓	✓

e.g., OCP[αlateral]), where the transvocalic versus unbounded distinction is encoded as a parameter on the relevant autosegmental rule (Odden, 1994; Piggott, 1996) or markedness constraint (Pulleyblank, 2003; Suzuki, 1998). Alternatively, if non-TSL$_2$ patterns are within the reach of human phonotactic learners (contra Lai, 2015), then we still have no reason to expect assimilation and dissimilation patterns to differ. In **Table 2**, this hypothesis is labelled L(T)T, referring to the more complex Locally Testable or Locally Threshold Testable regions of the subregular hierarchy, within which strictly beyond-transvocalic dependencies would fall. Finally, the ABC model predicts an asymmetry: Strictly beyond-transvocalic locality should be learnable for dissimilation but not assimilation.

1.4. Testing the hypotheses: Learning experiments

In evaluating the predictions laid out in **Table 2**, the cross-linguistic record is of limited help. The only reported case that approximates a strictly beyond-transvocalic pattern is liquid dissimilation in Sundanese (Bennett, 2015a; Cohn, 1992). Here the plural infix /-ar-/ surfaces as [-al-] in the presence of another [r] (9a), but not in the transvocalic configurations of (9b)–(9c).

(9) Sundanese: Alternations in plural infix /-ar-/
    a. /ar/ + /ŋumbara/ ŋ<a.**l**>um.ba.**r**a 'go abroad (pl.)'
    b. /ar/ + /curiga/ c<a.**r**>u.**r**i.ga 'suspicious (pl.)'
    c. /ar/ + /rahɨt/ **r**<a.**r**>a.hɨt 'wounded (pl.)'
    d. /ar/ + /hormat/ h<a.**l**>o**r**.mat 'respect (pl.)'
    e. /ar/ + /litik/ **l**<a.**l**>i.tik 'little (pl.)'
    f. /ar/ + /gilis/ g<a.**r**>i.**l**is 'beautiful (pl.)'

However, Sundanese is not entirely persuasive as evidence of strictly beyond-transvocalic dissimilation, due to a number of complicating factors. For one thing, dissimilation does apply within the transvocalic window, provided that the two liquids occupy different syllable constituents (onset versus coda), as in (9d). Secondly, within the transvocalic window there is also liquid *assimilation* under some conditions (9e) but not others (9f). Although Bennett (2015a) presents a complete analysis in ABC terms, other interpretations are possible. For example, Stanton (2018) argues, based on quantitative lexical data, that the full pattern reflects an essentially unbounded dissimilation of [r…r] sequences which is overridden by a pattern of *aggressive reduplication* (Zuraw, 2002) holding for pairs of adjacent syllables (especially word-initial $\sigma_1\sigma_2$ pairs).

We must therefore turn to other methods to investigate whether strictly beyond-transvocalic locality is learnable and whether dissimilation and assimilation differ in this respect. In situations like these, where the cross-linguistic typology is either too sparse to distinguish systematic gaps from statistical accidents or inherently ambiguous as to the source of those gaps (see Section 1.2.1), *artificial grammar learning* (AGL) experiments are a particularly helpful tool (Culbertson, 2012; Finley, 2017; Moreton & Pater, 2012a, 2012b).

Certain aspects of the question have been investigated using AGL methods. For sibilant harmony over the (English) [s]-[ʃ] distinction, Finley (2011, 2012) found that English adults seem to interpret interaction in beyond-transvocalic contexts as evidence of an unbounded dependency. When exposed to stems of the shape S̲VCV, CVS̲VCV, or S̲VCVCV (where S is [s] or [ʃ]), together with affixed forms containing a (harmonic) suffix element [-su] ~ [-ʃu], learners generalized the dependence of suffix choice on stem sibilant to unfamiliar trigger-target distances, and crucially to transvocalic configurations like (CV)CVS̲V-S̲u. McMullin and Hansson (2014) replicated this finding in a design where

two distinct suffixes ([-si], [-ʃu]) each triggered a harmony alternation in the preceding stem. Again, participants exposed to evidence of assimilation in beyond-transvocalic CVS̲VCV-S̲V contexts extended this to transvocalic CVCVS̲V-S̲V. By contrast, both studies found that participants exposed only to items with harmony in transvocalic (CV)CVS̲V-S̲V contexts did not generalize this to beyond-transvocalic distances. This fits the predictions of both the TSL$_2$ hypothesis and the ABC model, but does not distinguish between the two.

Below we report on four AGL experiments that extend this line of inquiry in two important ways. Firstly, we examine learning of co-occurrence patterns at differing locality levels for assimilation and dissimilation alike, allowing comparison of the two. This makes sibilant-sibilant interactions like those featured in previous studies less suitable; whereas sibilant harmony is by far the most frequently attested type of long-distance consonant assimilation (Hansson, 2010), analogous dissimilations (e.g., /ʃ…ʃ/ → [s…ʃ]) are virtually unattested. In the interest of cross-linguistic plausibility, we avoid this typological asymmetry (and any substantive biases that might contribute to it) by using co-occurrence patterns involving *liquids* [l, r]. As illustrated above (Section 1.3), liquid harmony and liquid dissimilation are both robustly attested.

Secondly, we go beyond previous studies in manipulating the richness of the training data that participants encounter. In Experiments 1 and 2, which target assimilation and dissimilation, respectively, we follow Finley (2011, 2012) in adopting a 'poverty of stimulus' paradigm (Finley & Badecker, 2009; Wilson, 2006). Certain critical configurations are withheld from the exposure data, and the test phase examines whether learners generalize the observed pattern to those novel contexts, e.g., from beyond-transvocalic to transvocalic distance. In Experiments 3 and 4, we adopt a 'rich-stimulus' paradigm, which directly probes the learnability of the sound pattern under dispute, such as a strictly beyond-transvocalic dependency (cf. Lai, 2015 on first-last harmony, White, 2014 on saltational alternations). Here the exposure data are augmented so as to make their full range compatible only with the pattern at stake (e.g., containing evidence not only of interaction in beyond-transvocalic contexts but also of *non*-interaction in transvocalic contexts). If learners fail to learn the target pattern even when provided with such 'rich' learning data, or if they go against the data by inferring a simpler pattern (e.g., an unbounded dependency), then this is especially persuasive evidence of a strong learning bias against the type of pattern in question.

The remainder of this paper proceeds as follows. In Section 2, we describe aspects of the methodology that apply to all four experiments. Section 3 then summarizes the training conditions for each of the experiments and lays out the full range of predictions under the ABC and TSL$_2$ hypotheses. The results of Experiments 1 and 2 are presented in Sections 4 and 5. After motivating the use of a 'rich-stimulus' training regime in Section 6, the results of Experiments 3 and 4 are presented in Sections 7 and 8. Finally, Section 9 discusses the overall findings and evaluates the competing hypotheses, and Section 10 concludes.

2. General methodology

2.1. Participants

Participants were recruited through a pool consisting of students at the University of British Columbia. Each participant could sign up for at most one of the experiments, and was compensated with either $10 or course credit. Although no participants were excluded based on language background, the results reported here only include data from self-reported native speakers of English. Though many of these were not monolingual, no participant had experience with any language known to exhibit a long-distance consonant interaction.

2.2. Stimuli

The entire set of stimuli, designed for use in Experiments 1 through 4, consisted of 1,560 items. This included 'verb stems,' as well as suffixed versions of each stem, where the suffixes [-li] and [-ɹu] correspond to 'future' and 'past' tense, respectively.[2] Note that the target pattern in each of the experiments is one with *regressive* directionality. The suffixes thus surface faithfully, triggering a pattern of liquid harmony or dissimilation that is exhibited in the form of stem allomorphy. As such, each of the training stems containing a liquid requires a corresponding set of four suffixed forms (each of the two stem allomorphs concatenated with each of the two suffixes). The breakdown of the types of stimuli in each phase of the experiment is shown in **Table 3**, and the full list is provided in the Appendix. These stimuli were divided into four lists of 390, each containing an equal number of each consonant and vowel in each position of the verb stems for each of the three phases (described below). Each list was randomized and recorded in a soundproof booth by one of four phonetically trained native speakers of North American English (two male, two female). While these speakers

Table 3: Summary of stimuli recorded for Experiments 1 through 4; L denotes a liquid [l, ɹ], C is one of [p, t, k, b, d, g, m, n], and V is a vowel [i, e, o, u].

Phase	Stem type (#)	Suffixed forms (#)	Total # of stimuli
Practice	CVCV (8)	CVCV-li (8)	24
		CVCV-ɹu (8)	
Training	CVCVLV (96)	CVCVlv-li (96)	480
		CVCVɹv-li (96)	
		CVCVlv-ɹu (96)	
		CVCVɹv-ɹu (96)	
	CVLVCV (96)	CVlVCV-li (96)	480
		CVɹVCV-li (96)	
		CVlVCV-ɹu (96)	
		CVɹVCV-ɹu (96)	
	CVCVCV (96)	CVCVCV-li (96)	288
		CVCVCV-ɹu (96)	
Testing	CVCVLV (32)	CVCVlv-li (16)	96
		CVCVɹv-li (16)	
		CVCVlv-ɹu (16)	
		CVCVɹv-ɹu (16)	
	CVLVCV (32)	CVlVCV-li (16)	96
		CVɹVCV-li (16)	
		CVlVCV-ɹu (16)	
		CVɹVCV-ɹu (16)	
	LVCVCV (32)	lVCVCV-li (16)	96
		ɹVCVCV-li (16)	
		lVCVCV-ɹu (16)	
		ɹVCVCV-ɹu (16)	
		Total = 1560	

---

[2] As the rhotic used in our stimuli was that of North American English, we will henceforth use the transcription [ɹ] whenever discussing the stimuli or the sound patterns exhibited across sets of stimuli.

were told that the stimuli would be used in an artificial grammar learning experiment, they were unaware of the sound patterns under investigation. Each speaker was instructed to produce the stimuli with initial stress, without vowel reduction, and with all segments as they would in normal speech.

2.3. Experimental design: Three phases

Experiments 1 through 4 each included three phases: practice (identical for all groups), training (differed by group), and testing (identical for all groups). Each experiment took approximately 45 minutes to complete.

Practice phase   A set of eight verb stems (along with their two suffixed forms) was constructed for the practice phase, in which participants learned how to conjugate the verbs of the artificial language in past and future tense. First, participants were told that they would be listening to pairs of words consisting of a present tense verb followed by its past tense form. They then listened, over headphones, to eight pairs of words consisting of a bare verb stem followed by its past tense form with the suffix [-ɹu] (e.g., [toke]… [toke-ɹu]). The interval between each stem and its suffixed form was 500 ms. The same procedure was then repeated for future tense forms (e.g., [toke]…[toke-li]). The purpose of this phase was to teach participants, prior to exposure to a phonological pattern, that the past and future tenses are formed by adding a suffix to the present tense form of the verb. Note that, in contrast to the training phase described below, participants were not asked to repeat these practice items out loud. Furthermore, to minimize any influence on the remainder of the experiment, all stems in the practice phase were disyllabic (CVCV) with no liquids.

Training phase   The training phase for each condition included a series of 192 triplets, each consisting of a trisyllabic stem followed by its two suffixed forms with [-li] and [-ɹu], all three produced by the same talker. This phase was self-paced, with participants using the keyboard to advance through the items. Participants were asked to repeat each word aloud after hearing it; this is known to aid learning in similar tasks (e.g., Warker, Xu, Dell, & Fisher, 2009). The precise contents of the training phase for each experimental condition are provided in the relevant sections below.

   One portion of the training stimuli shown in **Table 3** included 96 CVCVLV stems (where L is [l] or [ɹ]), to be used for the 'Short-range' training conditions. Note that since the list of stimuli was designed to be used for the full series of experiments, which cover both liquid harmony and dissimilation, four suffixed forms were recorded for each such stem. For example, a verb stem like [pidele] had two forms adhering to a suffix-triggered pattern of liquid harmony (future [pidele-li], past [pideɹe-ɹu]), and two that were instead consistent with suffix-triggered dissimilation ([pideɹe-li], [pidele-ɹu]). For each Short-range training stem, the list of recorded stimuli therefore included five words (one bare stem, four suffixed forms), resulting in 480 separate stimuli. These were counterbalanced for the number of stems with each liquid, which talker produced the stimuli for each set of stem + suffix items, and the frequency of the non-liquid segments in each position.

   Similarly, 96 'Medium-range' stems (CVLVCV) were recorded, along with each of their four possible suffixed forms (480 total). To maintain counterbalancing, each Medium-range stem was obtained by reversing the order of the second and third syllables of the Short-range items (e.g., Short-range [kopeɹu] corresponded to Medium-range [koɹupe]).

   The final subset of training stimuli consisted of 96 CVCVCV verb stems along with their suffixed forms. Since these stems contained no liquids, only two suffixed versions of each stem were required (e.g., [dutebi], [dutebi-li], [dutebi-ɹu]), resulting in 288 words.

Testing phase   Finally, the testing phase used a two-alternative forced choice (2AFC) paradigm to determine whether participants preferred liquid agreement (harmony) or disagreement (disharmony) at three levels of distance: Short-range (CVCVL̲V-L̲V), Medium-range (CVL̲VCV-L̲V), and Long-range (L̲VCVCV-L̲V). On each of 96 trials (32 for each distance), participants heard an unfamiliar verb stem containing a liquid, and were asked to choose the correct option from two possible suffixed forms, each with the same suffix. For example, one Long-range test item was [ɹomuge], with the options [lomuge-li] and [ɹomuge-li]. All three items in each trial were produced by the same talker. An interval of 500 ms separated the bare stem and the first suffixed option, with 250 ms between the two suffixed options. Participants had 3 seconds to respond after the onset of the second option before receiving a no-response error message. The set of testing phase stimuli (96 triplets = 288 total) was counterbalanced for several factors, including which liquid the stem contained, which suffix was used, which talker produced the stimuli, whether the first or second 2AFC option had liquid harmony, which 2AFC option involved an alternation (in relation to the bare stem), and the number of non-liquid consonants and vowels in each position. Example trials are shown in **Table 4**.

3. Summary of training conditions and predictions

3.1. Experiments 1 and 2

The purpose of Experiments 1 and 2 is to ensure that we can extend previous findings about sibilant harmony (Finley, 2011, 2012; McMullin & Hansson, 2014) to liquids, and to thereby verify that the current design is adequate for a direct comparison of long-distance consonant agreement and disagreement. Experiments 1 and 2 therefore use a 'poverty of stimulus' training paradigm, in which the training phase for each group includes direct evidence of the target dependency either at the Medium-range distance or the Short-range distance. However, participants are given no information as to the behaviour of co-occurring liquids at any other distance level. Throughout the remainder of this paper, we use the following labels to refer to each of these experimental groups: M-Harm (for Medium-range harmony), S-Harm (Short-range harmony), M-Diss (Medium-range dissimilation), and S-Diss (Short-range dissimilation). A summary of the types of items that were presented to and withheld from each group during training is presented in **Table 5**.

Experiment 1 predictions   Based on the findings of Finley (2011, 2012), the M-Harm group is predicted to learn an unbounded pattern that applies at each of the three testing distances: Short-range, Medium-range, and Long-range, whereas the S-Harm group is predicted to restrict the pattern to Short-range contexts (even though their training data is in principle consistent with an unbounded pattern as well). These results would be compatible with both the ABC and TSL_2 hypotheses as laid out in Section 1.3.

Table 4: Example of testing trials for all experiments.

Distance	Stem	2AFC options
Short-range	pidole	pidole-ɹu … pidoɹe-ɹu
(CVCVL̲V-L̲V)	duteɹe	dutele-li … duteɹe-li
Medium-range	tuluge	tuɹuge-li … tuluge-li
(CVL̲VCV-L̲V)	miɹete	milete-ɹu … miɹete-ɹu
Long-range	lugoni	lugoni-ɹu … ɹugoni-ɹu
(L̲VCVCV-L̲V)	ɹomuge	lomuge-li … ɹomuge-li

Table 5: Training conditions for Experiments 1 and 2. Cells marked with '✓' indicate that evidence of the target dependency at that distance was presented in the training phase. Cells marked with '?' indicate that no exposure to liquid co-occurrences at this distance was provided during training. (Participants were tested at all three distances).

Exp.	Group	CVCV$\underline{L}$V-$\underline{L}$V	CV$\underline{L}$VCV-$\underline{L}$V	$\underline{L}$VCVCV-$\underline{L}$V
1	M-Harm	?	✓	?
	S-Harm	✓	?	?
2	M-Diss	?	✓	?
	S-Diss	✓	?	?

Experiment 2 predictions   In the ABC model, the same constraints that help generate strictly transvocalic harmony will, given certain ranking configurations, generate *strictly beyond-transvocalic* dissimilation, as illustrated in (3)–(5) and (6)–(8) above. By the ABC hypothesis, a strictly beyond-transvocalic dependency is therefore within reach of the learner, but only if the pattern is a *dissimilatory* one. If learners are reluctant to generalize to contexts not encountered in training (unless such generalization is the only option available in the hypothesis space), then we should expect the M-Diss group to infer that the pattern is strictly confined to beyond-transvocalic contexts (Medium- and Long-range test items, but not Short-range ones). If M-Diss participants instead infer an unbounded dependency (generalizing to Short- and Long-range items), then such a result could only be reconciled with the ABC hypothesis through additional assumptions about how learners traverse the space of available grammars, e.g., through built-in priors on the relative ranking/weighting of constraints (we take up this question in Section 9.2.1 below).

The results of the S-Diss group do not bear on our comparison of the TSL_2 and ABC hypotheses. However, as noted in Section 1.3, the ABC literature employs two different types of locality-based constraints: those that penalize correspondence in non-adjacent syllables (PROXIMITY/CC-SYLLADJ) and those that enforce correspondence only in CVC contexts (e.g., CORR-LIQ$_{CVC}$). If the S-Diss group does not generalize to longer distances, this could be taken as support for the inclusion of the latter constraint type, as these are required to generate a strictly transvocalic dissimilation pattern within the ABC model.

By contrast, there is no substantive difference between TSL_2 characterizations of agreement (harmony) and disagreement (dissimilation). From this perspective, participants in the M-Diss and S-Diss conditions are predicted to behave the same way as those in the M-Harm and S-Harm conditions of Experiment 1, respectively: The M-Diss group should infer an unbounded pattern, generalizing to Short- and Long-range contexts, and the S-Diss group should learn a strictly transvocalic pattern, failing to generalize beyond the Short-range context.

3.2. Experiments 3 and 4

The purpose of Experiments 3 and 4 is to determine whether learners can detect strictly beyond-transvocalic dependencies when their training exhibits unambiguous evidence of such a pattern. To achieve this, we implement a 'rich stimulus' paradigm, augmenting the training data with direct evidence that the target pattern is not enforced at all distances.

The four new experimental groups are labelled M-Harm-S-Faith, S-Harm-M-Faith, M-Diss-S-Faith, and S-Diss-M-Faith. These labels follow the conventions established for previous experiments: 'S' indicates Short-range (CVCV$\underline{L}$V stems), 'M' indicates Medium-range (CV$\underline{L}$VCV stems), 'Harm' indicates that the target pattern at that distance was harmony, and

'Diss' indicates a target pattern of dissimilation. To further differentiate these groups from those in Experiments 1 and 2, the distance that showed overt evidence of *non*-alternation (faithful preservation of the stem liquid in the suffixed forms) is indicated by 'S-Faith' or 'M-Faith' at the end of each group's label. A summary of the types of training items included for each of the training groups is given in **Table 6**.

Experiment 3 predictions   Both the ABC and TSL$_2$ accounts of consonant harmony allow for transvocalic and unbounded locality, but neither can generate a strictly beyond-transvocalic assimilation pattern. As such, they do not differ with respect to their predictions about Experiment 3. Specifically, the training data for the S-Harm-M-Faith group is consistent with strictly transvocalic harmony, and is therefore expected to be learned as-is. However, since the training data for the M-Harm-S-Faith group is not compatible with unbounded or strictly transvocalic locality, both the TSL$_2$ and ABC hypotheses predict that participants will deviate from the training pattern during the test phase. This expected bias against strictly beyond-transvocalic locality could manifest itself in one of two ways. First, participants could force a pattern of liquid harmony into the TSL$_2$/ABC hypothesis space by learning an unbounded pattern of liquid harmony (over-generalizing into Short-range contexts). Alternatively, if the pattern presented in training is simply too complex for participants to detect any systematic dependency, they should behave indistinguishably from the Control group, either selecting the faithful (non-alternation) response option throughout the test phase, for all three distances, or selecting items at random that illustrate liquid alternations.

Experiment 4 predictions   Note that the training data in the M-Diss-S-Faith condition adheres to a pattern of strictly beyond-transvocalic dissimilation. Moreover, since disagreement is not enforced in transvocalic contexts, the training pattern *cannot* be interpreted as an unbounded dependency. The ABC and TSL$_2$ hypotheses therefore differ from each other in their predictions for Experiment 4.

   Since strictly beyond-transvocalic dissimilation can be generated in the ABC model (see Section 1.3), participants in the M-Diss-S-Faith condition are predicted to be able to learn the pattern as presented in training. Conversely, all dependencies with strictly beyond-transvocalic locality (agreement and disagreement) fall outside of the TSL$_2$ region, suggesting that they should be inaccessible to learners. The TSL$_2$ hypothesis therefore predicts that participants in the M-Diss-S-Faith group will fail to learn the pattern exhibited in training, either over-generalizing the dependency to Short-range testing items or learning no discernible pattern whatsoever (as is predicted under both hypotheses for the M-Harm-S-Faith group in Experiment 3; see above).

Table 6: Training conditions for Experiments 3 and 4. Cells marked with '✓' indicate that evidence of the target dependency at that distance was presented in the training phase. Cells marked with '✗' indicate that liquid pairs at that distance, though encountered in training, did not conform to any systematic harmony or dissimilation pattern (instead staying faithful). Cells marked with '?' indicate that no exposure to liquid pairs at this distance was provided during training. (Participants were tested at all three distances).

Exp.	Group	CVCV<u>L</u>V-<u>LV</u>	CV<u>L</u>VCV-<u>LV</u>	<u>L</u>VCVCV-<u>LV</u>
3	M–Harm–S–Faith	✗	✓	?
	S–Harm–M–Faith	✓	✗	?
4	M–Diss–S–Faith	✗	✓	?
	S–Diss–M–Faith	✓	✗	?

4. Experiment 1: Liquid harmony (poverty of stimulus)

4.1. Participants

Fifty-five participants (42 female, 13 male, mean age 24.4) took part in Experiment 1. The first 48 of these were distributed across the three training conditions described below, 16 in each group. The remaining seven were recruited in order to obtain a minimum number of 'successful learners' in each experimental group, as described in Section 4.3.2.

4.2. Training conditions

Participants were assigned to one of three groups that differed in the type of verb stems encountered in training: two experimental groups (M-Harm and S-Harm), and a Control group. Recall that the training phase consisted of 192 triplets (a trisyllabic verb stem, followed by its two suffixed forms with [-li] and [-ɹu]).

For the experimental groups, half of the stems were distractors containing no liquids (e.g., [dutebi], [dutebi-ɹu], [dutebi-li]). The remaining half of the training stems were of the form CVLVCV for the M-Harm group (where L represents a liquid [l] or [ɹ]) but CVCVLV for the S-Harm group. Participants were exposed to a liquid harmony either across an intervening VCV sequence (M-Harm group; e.g., [piléde], [piléde-li], [piɹede-ɹu]) or across a single vowel (S-Harm group; e.g., [gutoɹo], [gutolo-li], [gutoɹo-ɹu]). For each of these stems, one of the two suffixes triggered an alternation resulting in harmony, whereas the other suffixed form obeyed harmony by morpheme concatenation alone. Training triplets were divided into two blocks, and randomized within block for each participant. Examples of training items for the experimental groups are provided in **Table 7**.

Participants in the Control group completed the same amount of training, but were not exposed to any stems containing liquids, and therefore saw no evidence for or against liquid harmony. Instead, in each block they encountered the full set of 96 triplets with non-liquid CVCVCV stems (and were thus exposed to each triplet twice). The Control group serves as a baseline for comparison in the statistical analyses, to help eliminate confounding effects of the experimental design, stimuli, or L1 influence (e.g., from English lexical statistics).

Table 7: Examples of training items for M-Harm and S-Harm groups in Experiment 1.

Group	Training triplet	Type and number of items
M-Harm	dutebi … dutebi-ɹu … dutebi-li	96 distractor stems
	mekotu … mekotu-li … mekotu-ɹu	
	pilede … pilede-li … piɹede-ɹu	48 stems with [l]
	neɹogi … neɹogi-ɹu … nelogi-li	
	koɹupe … kolupe-li … koɹupe-ɹu	48 stems with [ɹ]
	guɹoto … guɹoto-ɹu … guloto-li	
S-Harm	dutebi … dutebi-ɹu … dutebi-li	96 distractor stems
	mekotu … mekotu-li … mekotu-ɹu	
	pidele … pidele-li … pideɹe-ɹu	48 stems with [l]
	negiɹo … negiɹo-ɹu … negilo-li	
	kopeɹu … kopelu-li … kopeɹu-ɹu	48 stems with [ɹ]
	gutoɹo … gutoɹo-ɹu … gutolo-li	

4.3. Results and analysis

We present two statistical analyses of the Experiment 1 results. There is precedent for both approaches, and each has advantages and drawbacks. First, in Section 4.3.1, we analyze the responses of a predetermined number of participants (the first 16 in each group), without regard for whether these individuals showed evidence of learning the patterns they were exposed to. This method, common in AGL experiments using a poverty-of-stimulus paradigm (e.g., Finley, 2011, 2012; Finley & Badecker, 2009; Lai, 2015; Wilson, 2006), ensures that the data for each group are sampled from the same population. However, it somewhat obscures the question of interest: whether learners *generalize* from the types of contexts they encountered in training to novel contexts (e.g., for the M-Harm group, from Medium-range CVLVCV-LV to Short-range CVCVLV-LV). The very notion of 'generalization' presupposes that the exposure pattern has been learned in the first place, but this is not a given. Rather, individual performance often tends toward a bimodal distribution between successful learners and non-learners, instead of clustering around the group mean.

To address this problem, non-learners can be filtered out based on some threshold criterion for successful learning. Some studies (e.g., White, 2014) include a screening phase, such that only those who surpass the threshold continue to the testing phase; participants may be cycled repeatedly through the training and screening phases until they meet the criterion (within some time limit). While this method has the benefit of ensuring that each participant is adequately trained, different participants receive different amounts of training. This introduces a potential confound, e.g., if participants in one condition require (and hence receive) more training than those in another condition. Furthermore, as the intermittent screening phases involve 2AFC pairings of a 'correct' and an 'incorrect' form, they inevitably expose participants to ungrammatical data prior to the testing phase.

We instead examine the testing-phase responses from each participant post hoc, and classify as *successful learners* those who reliably instantiated a pattern of liquid harmony at the distance at which L…L sequences were present in their training data (Short-range for S-Harm, Medium-range for M-Harm). This approach allows us to investigate group-level effects among participants who all received the exact same amount of training. A predetermined number of successful learners (12 in each experimental group) was set as a target, and additional participants were recruited until that target was reached. In Section 4.3.2 we analyze the data from the first 12 successful learners in each experimental group, as compared against the first 12 participants in the Control group. This method ensures that all participants received an equal amount of training exposure; however, we cannot be certain that the results from each group truly represent the same population. We discuss these issues in Section 4.4.

4.3.1. Analysis of first 16 participants

A mixed-effects logistic regression model was fitted to the responses from the first 16 participants in each group of Experiment 1 (48 participants; 36 female, 12 male, mean age 23), using the *glmer* function in the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) of *R* (R Core Team, 2017).[3] The categorical dependent variable is whether, on a particular trial, the participant chose the response item that exhibited liquid harmony. The fixed-effects component, summarized in **Table 8**, includes as main effects the between-participants variable of Group (S-Harm and M-Harm, compared to Control) and the within-participants variable of trigger-target Distance in the test item (Medium- and Long-range, compared

---

[3] Preliminary results from Experiment 1, with 12 participants in each group, were reported in McMullin and Hansson (2014).

Table 8: Fixed-effects component of the logit mixed model for Experiment 1 (N = 4518; log-likelihood = −2184.2).

Coefficient	Estimate	SE	Pr (>|z|)
Intercept	0.43987	0.12132	0.0003
Harmony Second	−0.51084	0.14367	0.0004
Harmony Faithful	2.61950	0.30531	<0.0001
Medium-range	0.08113	0.18412	0.6595
Long-range	−0.16051	0.25510	0.5292
S-Harm	1.30893	0.38182	0.0006
M-Harm	1.25831	0.38503	0.0011
Medium-range × S-Harm	−1.00769	0.23162	<0.0001
Long-range × S-Harm	−0.95505	0.34226	0.0053
Medium-range × M-Harm	0.28217	0.23734	0.2345
Long-range × M-Harm	−0.38150	0.34577	0.2699

to Short-range), as well as an interaction between Group and Distance. Also included are two nuisance variables that significantly improved model fit: Harmony Second (whether the form with harmony was presented as the second of the two 2AFC items) and Harmony Faithful (whether the form with harmony involved mere morpheme concatenation, as opposed to an alternation in the stem liquid). The structure of the random-effects component was determined by the procedure recommended in Bates, Kliegl, Vasishth, and Baayen (2015), and included: random by-item and by-participant intercepts; random by-participant slopes for Harmony Second, Harmony Faithful, and Long-range distance; and random by-participant terms for the interactions between Long-range distance and each of Harmony Second and Harmony Faithful (as well as some correlation parameters). Full details are provided in the Appendix.

To reduce collinearity, all predictor variables were centered. As a result, the Intercept term reflects the grand mean across all groups, distances, etc.; overall, the odds for choosing the 2AFC option with harmony (matching liquids) were thus $e^{0.43987}$ = 1.55 to 1, translating into a probability of 1.55/(1 + 1.55) = 0.61. For each predictor, the coefficient estimate indicates the difference in log-odds between the two factor levels being compared. For instance, the odds of selecting the harmony option when that choice is presented second are only $e^{-0.51084}$ = 0.60 times those of doing so where harmony is presented first. The large estimate for Harmony Faithful indicates that participants are much more likely to select harmony on trials where that option involves a faithful stem liquid (e.g., [tuluge … tuɹuge-li, tuluge-li]) than where it involves an alternation (e.g., [pidole … pidoɹe-ɹu, pidole-ɹu]); the odds ratio is $e^{2.6195}$ = 13.73.

Because the model includes interaction terms for (the non-reference levels of) Group and Distance, the main-effect estimates for these predictors reflect the difference relative to the reference level of the factor in question (Control for Group, Short-range for Distance). For example, the non-significant effects for Medium- and Long-range show that the Control group is not reliably more or less likely to choose harmony on test trials involving these distances, as compared to trials with Short-range items. The significant main effects of S-Harm and M-Harm show that each experimental group is more likely than the Control group to choose harmony at the Short-range distance. The significant interactions of Medium- and Long-range distance with the S-Harm group show (with negative estimates) that this group applies harmony less often outside of the Short-range window than within

it. By contrast, neither of the interactions of Medium- and Long-range distances with this group approach significance, suggesting that the M-Harm group treats all three distances equally. A likelihood ratio test shows that inclusion of the Group × Distance interaction contributes significantly to the model fit ($\chi^2(4) = 37.02$, $p < 0.0001$).[4]

However, a comparison between distances within each experimental group does not answer the question of primary interest. Rather, we wish to determine, for each of the three distances, whether or not a given experimental group is significantly more likely than the Control group to choose harmony. This information can be obtained from the estimates for S-Harm and M-Harm in **Table 8**, but only for the baseline Short-range distance. We therefore re-fitted the same model, but with each of Medium-range and Long-range serving as the reference level for Distance, allowing for a direct comparison of groups at those distances as well. The combined results are presented in **Table 9**. Each cell shows an odds ratio (OR); for example, the (estimated) odds that a S-Harm group member would choose the harmony-obeying option for a Short-range test item (CVCVLV-LV) are more than three times (3.70) those of a Control participant doing so. The odds ratios in **Table 9** were obtained by exponentiating the coefficient estimates for S-Harm and M-Harm in the relevant version of the model. Thus, for example, the OR values in the Short-range column correspond to the estimates reported in **Table 8** ($e^{1.30893} = 3.70$, $e^{1.25831} = 3.52$).

The Experiment 1 results are displayed visually in **Figure 1**, which depicts the mean proportion of harmony responses for each group at each testing distance, as well as the proportions for individual participants. The plot and odds-ratio table both illustrate that the experimental groups showed evidence of *learning* the harmony pattern at the trigger-target distance encountered in training (dashed-line boxes in **Figure 1**, shaded cells in **Table 9**). Compared to the Control group, the M-Harm and S-Harm groups displayed significantly greater odds (4.67 and 3.70 times, respectively) of choosing the harmony response on test items of the relevant type.

As for *generalizing* that pattern to novel contexts, the M-Harm group chose harmony significantly more often than the Control group at the unfamiliar Short-range (OR = 3.52) and Long-range (OR = 2.40) distances. This indicates that the M-Harm group tends to interpret the pattern as an *unbounded* liquid harmony, generalizing from an observed assimilation in Medium-range CVLVCV-LV contexts both inwards to Short-range CVCVLV-LV and outwards to Long-range LVCVCV-LV contexts. By contrast, the S-Harm group does not generalize the pattern of liquid harmony from the Short-range distance to Medium-range (OR = 1.35, $p = 0.430$) or Long-range (OR = 1.42, $p = 0.106$) contexts.

Table 9: Experiment 1: Odds ratios for choosing harmony at each testing distance, comparing experimental groups to Control group (first 16 participants in each). Cells corresponding to contexts encountered in training are shaded; significance is indicated with boldface.

	Type of test item (trigger-target distance)		
	Short-range (CVCVLV-LV)	Medium-range (CVLVCV-LV)	Long-range (LVCVCV-LV)
M-Harm versus Control	**3.52**	**4.67**	**2.40**
	p = 0.001	**p < 0.001**	**p < 0.001**
S-Harm versus Control	**3.70**	1.35	1.42
	p < 0.001	p = 0.430	p = 0.106

---

[4] The same is true for each of the nuisance variables, Harmony Second and Harmony Faithful; the relevant statistics are not shown here.
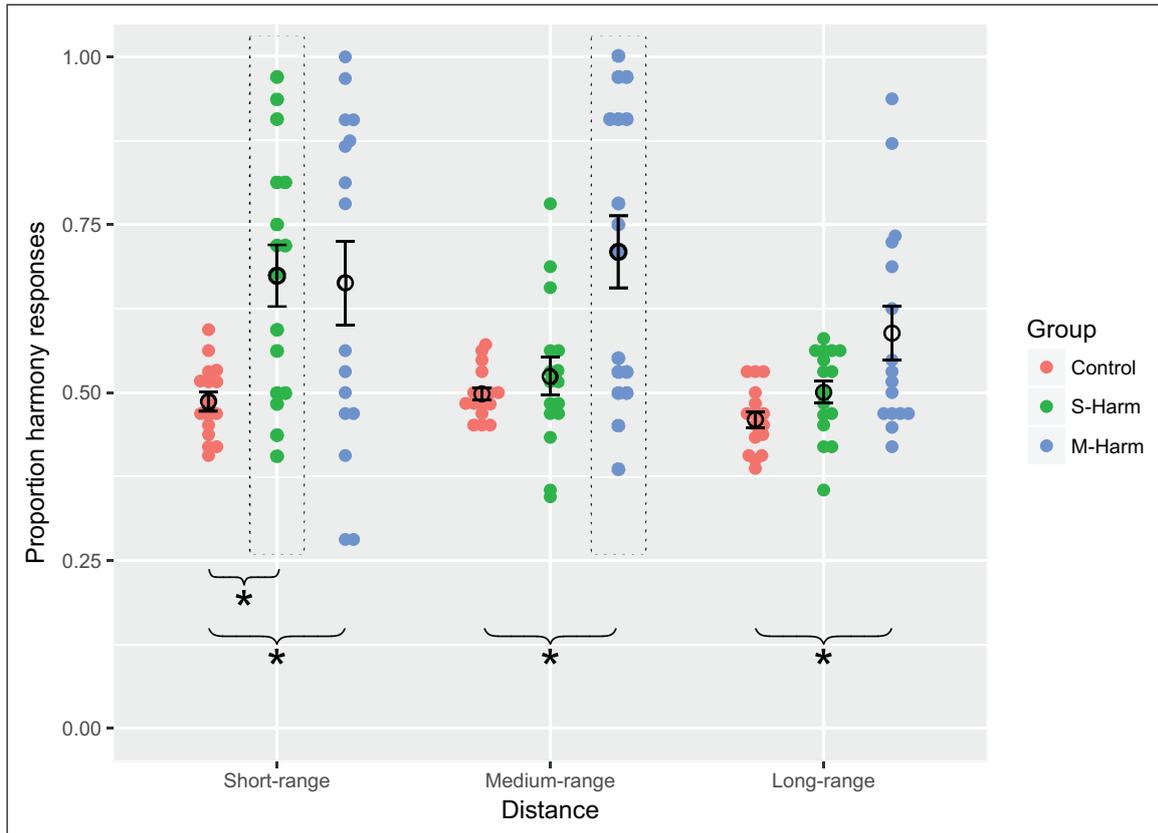
Figure 1: Proportions of harmony responses across conditions (groups) and test-item type (distance) for Experiment 1. Coloured dots show individual participant performance, with group means indicated by an open circle with error bars (showing standard error); significant between-group differences are extracted from the logit mixed model. Performance on contexts encountered in training is shown with dashed-line boxes.

4.3.2. Analysis of successful learners

The following procedure was used to obtain 12 'successful learners' in each of the S-Harm and M-Harm groups. The threshold for successful learning was defined as exhibiting a proportion of responses conforming to the exposure pattern (here, liquid harmony) *at the distance encountered in training* that surpassed the 95% confidence level on a one-tailed binomial test.[5] This threshold was defined independently of the previously collected data—that is, without knowing how many of the first 16 participants in each condition might qualify as learners. For example, if an M-Harm participant registered a response on all 32 of the relevant test items (Medium-range in this case), at least 21 of these would need to be harmony responses for the participant to count as a successful learner. Participants who did not register at least 29 responses on the relevant trials were not considered for this analysis. In terms of proportions of harmony responses, the threshold criterion applied varies slightly across participants, ranging from 63.3% (19/30) to 65.6% (21/32) depending on the total number of registered responses on the relevant trials.[6]

---

[5] A one-tailed test was used since successful learners are expected to choose items adhering to the exposure pattern more often than the Control group, rather than simply differing from them in either direction.

[6] While a success rate in the vicinity of 65% may at first glance seem a rather lax criterion, it is both justified and necessitated by the specifics of our stimuli and experiment design. First, studies which use a higher threshold (e.g., 80% in White, 2014) apply this to responses on previously encountered (recall) items. In our design, by contrast, all test items are novel; the relevant items are 'familiar' only in the sense that they contain a liquid in the same position in the stem as did the (non-distractor) items within the training data. In studies in which the testing phase includes 'old' (recall) items as well as novel items of the familiar type (e.g., Finley, 2011, 2012), application of the learned pattern is always markedly lower on the latter.

For reference, nine of the first 16 M-Harm participants met this criterion on Medium-range test items, as did nine of the first 16 S-Harm participants on Short-range test items. In order to obtain 12 successful learners in each of the S-Harm and M-Harm groups, an additional four and three participants, respectively, were recruited. Unsurprisingly, none of the 16 participants in the Control group surpassed the threshold at any of the three testing distances. To maintain consistency with the other groups, the analysis below includes only the first 12 Control participants, rather than the full set of 16.

The responses from these successful learners were analyzed by fitting a mixed-effects logistic regression model with structure identical to that used in the above analysis of the first 16 participants in each group. **Table 10** presents the odds ratios extracted from the fitted model; for full model details, see the Appendix.

As before, the M-Harm group is significantly more likely than the Control group to choose liquid harmony at all three levels of locality, generalizing from the Medium-range context to the novel Short- and Long-range contexts. By contrast, the S-Harm learners again do not significantly favour harmony outside of the Short-range context they encountered in training. However, both here and in **Table 9** the odds ratio estimates for the S-Harm versus Control comparison on the Medium- and Long-range items are above 1.0, even if not significantly so, suggesting that a few S-Harm participants may be generalizing harmony to those distances. We discuss such individual differences below.

4.4. Discussion of Experiment 1

To summarize, both analyses provide evidence that participants who are exposed to harmony in Medium-range contexts (spanning an intervening VCV sequence) tend to learn an *unbounded* dependency, generalizing both inwards to Short-range and outwards to Long-range contexts. By contrast, those exposed to consonant harmony in Short-range contexts (where no consonant intervenes) tend to interpret this conservatively: as a harmony

Table 10: Odds ratios for choosing harmony at each testing distance, comparing first 12 successful learners in each experimental group to first 12 participants in Control group. Cells corresponding to contexts encountered in training are shaded; significance is indicated with boldface.

	Type of test item (trigger–target distance)		
	Short-range (CVCVLV-LV)	**Medium-range (CVLVCV-LV)**	**Long-range (LVCVCV-LV)**
M-Harm versus~Control	**12.89**	**14.62**	**4.37**
	***p* < 0.001**	***p* < 0.001**	***p* < 0.001**
S-Harm versus~Control	**13.47**	1.70	1.50
	***p* < 0.001**	*p* = 0.071	*p* = 0.137

Furthermore, such higher threshold criteria (defined over recall items) are typically applied as part of a design that uses a pre-test screening phase; for example, many participants in White (2014) only reach 'success' (i.e., 80% or higher) after cycling through the exposure phase + screening phase multiple times. In our design, the threshold criterion is instead applied to performance on the testing phase itself. (See the introduction to Section 4.3 for our reasons not to use a screening phase and repeated exposure phases.) Secondly, our design differs from previous studies (e.g., Finley, 2011, 2012) in that the target pattern involves alternation within the open-class morphemes (stems) themselves, rather than mere concatenation of stems with one of two (previously-encountered) suffix allomorphs. This greatly increases the pressure toward levelling (faithful non-alternation), and hence lowers the overall proportion of 'successes' that can be realistically expected. Finally, although the criterion we use is defined based on a (one-tailed) binomial test, it should be noted that the variance of the Control group is considerably lower than what would be expected from a genuine binomial distribution. Thus a proportion of harmony responses that exceeds the threshold in fact represents an even more decisive departure from the Control group's performance than the characterization in terms of a 95% binomial confidence limit suggests.

requirement that targets only *transvocalic* consonant pairs. These findings conform to the typological dichotomy described in Section 1.1 above. They also replicate the findings of Finley (2011, 2012) and McMullin and Hansson (2014, Experiment 1), extending these to a different class of segments: liquids [l, ɹ] rather than sibilants [s, ʃ].

With respect to the differing predictions summarized in Section 3, these results are consistent with the TSL$_2$ hypothesis as well as the ABC hypothesis. Both incorporate the implicational universal that an assimilatory dependency at some beyond-transvocalic distance necessarily entails that the dependency is unbounded, holding at shorter and longer distances as well. Under the ABC and TSL$_2$ hypotheses, the training data encountered by the M-Harm group should be unambiguous evidence of unbounded harmony, while those seen by the S-Harm group are inherently ambiguous, equally compatible with a transvocalic or an unbounded dependency. Under the hypothesis that non-TSL$_2$ sound patterns in the Locally Testable (LT) or Locally Threshold Testable (LTT) regions are also within the range of human learners, the results of Experiment 1 (and those of the previous studies cited above) are unexpected. By the L(T)T hypothesis, the M-Harm training data should be no less ambiguous than the S-Harm data: Just as the latter could represent either a strictly-transvocalic pattern or an unbounded one, so could the former be evidence of either a strictly-*beyond*-transvocalic or an unbounded dependency. The fact that the S-Harm group tends to interpret their ambiguous training input conservatively, favouring a strictly-transvocalic interpretation, makes it all the more surprising that the M-Harm group behaves in the exact opposite way, favouring the less conservative unbounded interpretation over a strictly-beyond-transvocalic one. In summary, Experiment 1 provides strong support for the ABC and TSL$_2$ hypotheses over the L(T)T hypothesis, but does not distinguish between the former two.

Two final points are worth making regarding finer-grained aspects of the results. First, looking at the successful M-Harm learners, **Table 10** shows a dramatic difference in the effect size for generalizing harmony to Short-range (OR = 12.89) versus Long-range (OR = 4.37) contexts. (The same trend, though more subtle, is visible in **Table 9**). This is likely due to an independent reluctance to extend a phonological alternation into the highly salient word-initial position involved in the Long-range LVCVCV-LV test items (cf. Becker, Nevins, & Levine, 2012; the same effect was found in Finley, 2012). Recall in this connection that all stimuli were produced with initial-syllable prominence, contributing further to the salience of the target liquids in these test items.

Secondly, individual participants sometimes deviate from their group trend; this is particularly notable for the S-Harm condition. As mentioned above, the odds ratio estimates in **Tables 9** and **10** for the S-Harm versus Control comparisons on Medium- and Long-range test items suggest that some S-Harm participants are extending harmony to these longer distances, indicating that they inferred an unbounded pattern. Closer examination supports this interpretation: Three of the 12 successful S-Harm learners surpassed the threshold criterion on Medium-range items as well. None were above threshold on Long-range items, however, presumably due to the aforementioned general bias against extending alternation to a word-initial consonant. These observations are consistent with the hypothesis (see Section 1.2.1) that unbounded consonant harmony emerges diachronically as a secondary development through overgeneralization of an originally transvocalic pattern (Dolbey & Hansson, 1999).

5. Experiment 2: Liquid dissimilation (poverty of stimulus)

While the ABC and TSL$_2$ hypotheses converge in their predictions about possible locality patterns for long-distance assimilation (harmony), they diverge on dissimilatory dependencies. Specifically, patterns of disagreement with strictly beyond-transvocalic locality can be generated within the factorial typology of ABC constraints, but cannot be characterized

as TSL$_2$ formal languages (see **Table 2**). To test whether this predicted asymmetry is borne out in artificial grammar learning, we conducted an experiment analogous to Experiment 1 but for dissimilatory liquid-liquid interactions. The methodological details are nearly identical to those of Experiment 1, with the sole difference being the type of dependency that the experimental groups were exposed to during training.

5.1. Participants

Participants for two additional training conditions, M-Diss and S-Diss, were recruited and compensated as described above, yielding a total of 40 new participants for Experiment 2 (30 female, 10 male, mean age 24.1). The first 32 of these were distributed evenly across these two conditions, 16 in each. The additional eight were recruited in order to obtain 12 successful learners in each condition, just as in Experiment 1; see Section 5.3.2 for details.

5.2. Training conditions

All training stems were identical to those used in Experiment 1 (see **Table 7** above). However, instead of liquid harmony, participants in the experimental groups were exposed to consistent liquid *dissimilation* in the relevant suffixed forms. The M-Diss group was exposed to 96 triplets that exhibited dissimilation in CVL̲VCV-L̲V contexts (e.g., [pil̲ede], [pil̲ede-li], [pil̲ede-ɹu]), while the S-Diss group saw 96 corresponding triplets with CVCVL̲V-L̲V. As before, the training phase for each group also included 96 stems with no liquid. The testing phase in Experiment 2 was identical to that of Experiment 1 (see **Table 4** for examples).

5.3. Results and analysis

The Experiment 2 results are analyzed in the same way as those in Experiment 1. As the two experiments were designed and implemented in parallel, we use the response data from the same set of Control participants as in Experiment 1. An analysis of the first 16 participants in each group is presented in Section 5.3.1, followed in Section 5.3.2 by an analysis of the first 12 successful learners obtained in each group.

5.3.1. Analysis of first 16 participants

A mixed-effects logistic regression model was fitted to the responses of the first 16 participants in each condition (48 in total), with choice of the response option with two *different* liquids as the dependent variable. As with Experiment 1, the fixed-effects component includes the predictors of Group (M-Diss and S-Diss, compared against Control) and Distance (Medium- and Long-range, compared against Short-range), and an interaction between Group and Distance. The model again includes two nuisance variables, Dissimilation Faithful (whether the stem liquid in the option with disagreeing liquids was faithful to that of the bare-stem form) and Dissimilation Second (whether the option with disagreeing liquids was presented as the second 2AFC item). The random-effects structure was selected by the same procedure described for Experiment 1, and consists of random by-item and by-participant intercepts, random by-participant slopes for Dissimilation Second, Dissimilation Faithful, and Distance (Medium- and Long-range), and random by-participant terms for the interactions between Dissimilation Faithful and each of Medium- and Long-range distance (as well as a correlation parameter for these two interaction terms).
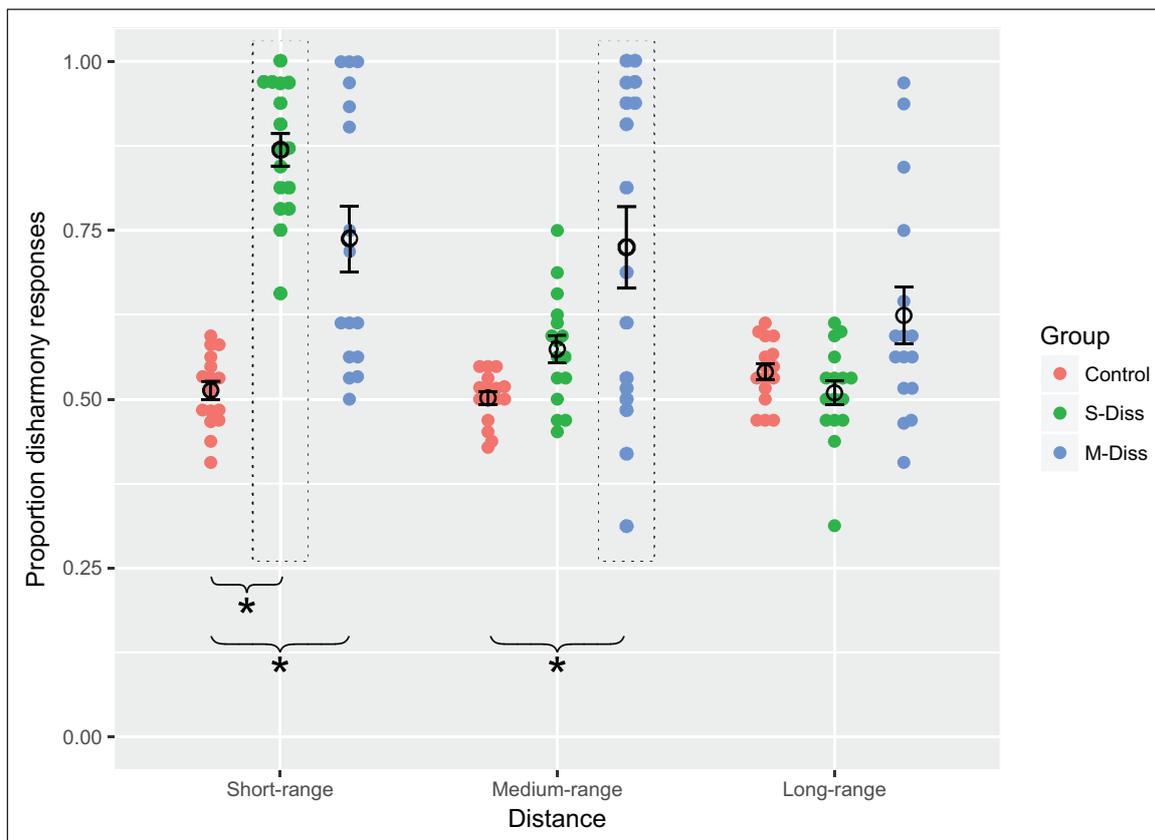
A likelihood ratio test shows that the Group × Distance interaction significantly improves model fit ($\chi^2(4) = 79.57$, $p < 0.0001$).[7] Full details of the model, including coefficient

---

[7] As in the case of Experiment 1, the same is also true for each of the nuisance variables, Dissimilation Faithful and Dissimilation Second.

estimates, are provided in the Appendix. Here we instead focus on the more informative odds-ratio comparisons between the experimental and Control groups at each test-item distance. As before, odds ratios are extracted from the model by exponentiating the coefficient estimates for the relevant Group predictor (M-Harm and S-Harm, respectively); to obtain such estimates for Medium-range and Long-range, the same model was re-fitted using each of these as the reference level of the Distance predictor (instead of Short-range). This comparison of each experimental group's performance against that of the Control group is provided in **Table 11**. A visual representation of the results from Experiment 2 is provided in **Figure 2**.

Table 11: Experiment 2: Odds ratios for choosing harmony at each testing distance, comparing experimental groups to Control group (first 16 participants in each). Cells corresponding to contexts encountered in training are shaded; significance is indicated with boldface.

	Type of test item (trigger-target distance)		
	Short-range (CVC$\underline{L}$V-$\underline{L}$V)	**Medium-range (CV$\underline{L}$VCV-$\underline{L}$V)**	**Long-range ($\underline{L}$VCVCV-$\underline{L}$V)**
M-Diss versus Control	**4.19**	**4.08**	1.48
	(p < 0.001)	**(p < 0.001)**	p = 0.218
S-Diss versus Control	**10.08**	1.57	0.78
	(p < 0.001)	p = 0.150	p = 0.423

Figure 2: Proportions of disharmony responses across conditions (groups) and test-item type (distance) for Experiment 2. Coloured dots show individual participant performance, with group means indicated by an open circle with error bars (showing standard error); significant between-group differences are extracted from the logit mixed model. Performance on contexts encountered in training is shown with dashed-line boxes.

Both experimental groups show evidence of learning. Overall, the M-Diss group enforces dissimilation in Medium-range contexts; they are more than four times (OR = 4.08) more likely than the Control group to choose dissimilation in CVLVCV-LV test items. Likewise, the S-Diss group is over ten times (OR = 10.08) more likely than the Control group to choose dissimilation in the CVCVLV-LV context that they encountered in training. The M-Diss group seems to generalize this pattern inwards to the Short-range test items (OR = 4.19); for Long-range (LVCVCV-LV) test items, the corresponding effect (OR = 1.48) is too small to reach significance. Finally, the S-Diss group does not display a significant preference for dissimilation in either Medium- or Long-range contexts.

5.3.2. Analysis of successful learners

To identify successful learners in each experimental group, we applied threshold criteria identical to those used in the Experiment 1 analysis (see Section 4.3.2), but defined in terms of the number of *disharmony* responses for the relevant test-item types. As for Experiment 1, the target number of successful learners was set at 12. To reach this target, an additional eight participants were recruited for the M-Diss training condition (nine of the original pool of 16 participants were classified as successful learners). The original group of 16 S-Diss participants already included 12 successful learners.

The results were analyzed by again fitting a logit mixed model with the same structure as in the analysis of the first 16 participants per group. **Table 12** presents the estimated odds ratios extracted from this model; see Appendix for full model details).

These results are largely in line with those obtained for the first 16 participants per group. Once again, the S-Diss learners do not reliably extend the liquid dissimilation (disharmony) that they apply in Short-range contexts to Medium- or Long-range contexts. For the M-Diss group, we see evidence of generalization of the learned Medium-range dependency both inward to Short-range (OR = 9.54) and outward to Long-range contexts (OR = 2.06). This differs slightly from the results for the first 16 participants per group (see **Table 11**), where the difference between M-Diss and Control groups was not significant on Long-range test items. There is however a considerable asymmetry in the effect sizes for the Short-range and Long-range contexts (cf. also Experiment 1).

5.4. Discussion of Experiment 2

The goal of Experiment 2 was to replicate Experiment 1 as faithfully as possible, but for dissimilatory as opposed to assimilatory dependencies. The purpose was to evaluate the competing TSL_2 and ABC hypotheses, with which the Experiment 1 results were equally compatible.

Table 12: Odds ratios for choosing disharmony at each testing distance, comparing first 12 successful learners in each experimental group to first 12 participants in Control group. Cells corresponding to contexts encountered in training are shaded; significance is indicated with boldface.

	Type of test item (trigger–target distance)		
	Short-range (CVCVLV-LV)	Medium-range (CVLVCV-LV)	Long-range (LVCVCV-LV)
M-Diss versus Control	**9.54**	**13.51**	**2.06**
	(p < 0.001)	**(p < 0.001)**	**p = 0.010**
S-Diss versus Control	**7.48**	1.36	0.81
	(p < 0.001)	p = 0.260	p = 0.460

The S-Diss group clearly preferred dissimilation in Short-range test items, indicating learning of the exposure pattern, but did not extend this dissimilation preference to either Medium- or Long-range contexts. They thus appear to have inferred a strictly transvocalic dissimilation dependency. This result is in line with the predictions of the TSL_2 hypothesis as well as those of the ABC model. It is also consistent with the cross-linguistic typology of consonant dissimilation; transvocalic dissimilation is relatively common and need not imply dissimilation at greater distances. It is noteworthy that, just like the S-Harm group in Experiment 1, S-Diss participants tended to interpret their inherently ambiguous training data conservatively, preferring a strictly transvocalic interpretation over an unbounded one.

As for the M-Diss condition, both analyses show this group to be significantly more likely to choose disagreeing liquids not only for Medium-range but also Short-range items. This is consistent with the TSL_2 hypothesis, by which interaction in beyond-transvocalic contexts entails interaction in transvocalic contexts as well (i.e., an unbounded dependency). It is somewhat unexpected under the ABC hypothesis, by contrast, as that model predicts strictly beyond-transvocalic dissimilation to be a perfectly possible (and hence learnable) sound pattern. Given that participants in the S-Harm and S-Diss conditions tended to act conservatively, we might similarly have expected the M-Diss group *not* to generalize the observed Medium-range dependency into Short-range contexts. Finally, both hypotheses treat the two non-syllable-adjacent distances (Medium- and Long-range) as equivalent and thus predict that if dissimilation is enforced at Medium-range the same should hold at Long-range. At first glance, the fact that the first 16 participants in the M-Diss group showed no significant preference for dissimilation on Long-range test items is thus somewhat problematic for both hypotheses. However, the analysis of the 12 successful learners suggests that those who learned dissimilation at the Medium-range distance did indeed tend to generalize outwards to Long-range contexts.

All in all, the Experiment 2 results reveal the same patterns of learning and generalization for dissimilation as Experiment 1 did for harmony. Crucially, the M-Diss group *did not* interpret dissimilation in Medium-range contexts as indicative of a *strictly beyond-transvocalic* dependency; this despite the fact that such a locality pattern is predicted by the ABC hypothesis to be possible, and that a strictly beyond-transvocalic (or perhaps strictly Medium-range) interpretation would have parallelled the 'conservatism' exhibited by the S-Diss (and S-Harm) group. Rather, the behaviour of the M-Diss group suggests that they tended to infer an unbounded dissimilation pattern, consistent with the predictions of the TSL_2 hypothesis—though obscured in Long-range test items by a general dispreference for alternation in word-initial syllables (just as in Experiment 1). In sum, the same dichotomy between strictly transvocalic and unbounded locality appears to emerge in the learning of dissimilation patterns as for harmony patterns.

6. Interlude: Poverty versus richness of the stimulus

Experiments 1 and 2 used a 'poverty of stimulus' paradigm (Finley & Badecker, 2009; Wilson, 2006) to determine not only whether English adults can learn assimilatory or dissimilatory dependencies between non-adjacent liquids, but also whether or not they generalize the observed pattern to contexts that are purposely withheld in training. The results suggested that participants who learn liquid harmony or dissimilation from Medium-range CVLVCV-LV items tend to generalize this to all levels of distance, indicating that they have learned an *unbounded* dependency. By contrast, participants who learn from Short-range CVCVLV-LV items tend to restrict the pattern to the *transvocalic* distance encountered in the training data. Importantly, neither of the two Medium-range training groups in Experiments 1 or 2 (M-Harm or M-Diss) showed evidence of learning a strictly *beyond-transvocalic* pattern (i.e., applying the pattern only to Medium- and Long-range test items, but not Short-range ones).

In particular, the fact that the M-Diss group did not infer a beyond-transvocalic pattern is problematic for the Agreement by Correspondence model, which predicts that strictly beyond-transvocalic *dissimilation* should be part of the learners' hypothesis space (while, notably, strictly beyond-transvocalic *harmony* should not; see Section 1.3).

However, we cannot conclude from these experiments that strictly beyond-transvocalic locality falls outside the hypothesis space of human phonological learners. Recall that the M-Harm and M-Diss groups were exposed to a dependency in Medium-range CV$\underline{\text{L}}$VCV-$\underline{\text{L}}$V contexts. Since the training phase provided no information about patterns of liquid-liquid co-occurrence in Short-range (CVCV$\underline{\text{L}}$V-$\underline{\text{L}}$V) or Long-range ($\underline{\text{L}}$VCVCV-$\underline{\text{L}}$V) contexts, the observed data were equally compatible with an unbounded interpretation as a strictly beyond-transvocalic one. The fact that participants tended to generalize the dependency to the transvocalic Short-range context does not entail that they are *incapable* of learning a strictly beyond-transvocalic dependency. Instead, the strongest conclusion we can draw is that, when presented with data that are ambiguous between unbounded and strictly beyond-transvocalic locality, learners have a strong *preference* for the former. Note that a preference for one analysis over another was also seen in the S-Harm and S-Diss groups, who evidently favoured a strictly transvocalic analysis over an unbounded one.

We therefore conducted two additional experiments, in which participants were instead provided with a 'rich stimulus' training phase. In these training conditions, the exposure data offered evidence about co-occurring liquids in Short-range and Medium-range contexts alike, rather than at just one of the two distances. The goal was to determine whether or not English adults can detect and learn a dependency that holds *exclusively* for beyond-transvocalic liquid-liquid pairs (represented by the Medium-range context).

7. Experiment 3: Locality-restricted harmony (rich stimulus)

Given that the ABC model predicts that patterns with strictly beyond-transvocalic locality should be learnable, but only for dissimilation (see **Table 2**), a question of particular interest is whether the results from Experiment 2 hold up even under a training regime where the exposure data provide overt evidence not only of dissimilation at a beyond-transvocalic distance but also of *non*-dissimilation (unrestricted co-occurrence, faithful non-alternation) in transvocalic contexts. The TSL$_2$ hypothesis, by contrast, predicts that strictly beyond-transvocalic dependencies should be unlearnable regardless of whether they involve harmony (agreement) or dissimilation (disagreement).

Even though both the ABC and TSL$_2$ hypotheses predict strictly beyond-transvocalic harmony to be impossible (in contrast to strictly transvocalic harmony), we start with examining learning of *harmony* patterns under this 'richness of the stimulus' paradigm, setting dissimilation aside for later (Experiment 4). This is motivated by the fact that a rich-stimulus training regime is qualitatively different from that of Experiments 1 and 2, which might create confounds when comparing results across experiments. Given that the predictions about possible versus impossible locality patterns for harmony are less controversial, and that the results from Experiment 1 were somewhat more consistent than those of Experiment 2 (in particular as regards generalization from Medium- to Long-range contexts), it is important to first replicate the Experiment 1 findings with a rich-stimulus training regime, before extending that design to dissimilation patterns.

7.1. Participants, Stimuli, and Procedure

Thirty-two participants (26 female, 6 male, mean age 23) were recruited and compensated as described for the previous experiments, with 16 assigned to each of the M-Harm-S-Faith and S-Harm-M-Faith conditions described below. The experiment was performed using the same stimulus set and procedures as for Experiments 1 and 2.

7.2. Training Conditions

The two new groups of participants were subjected to the exact same testing phase as in Experiments 1 and 2, but were exposed to a substantively different training phase. Recall that in Experiments 1 and 2, only 50% of the training stems for each of the experimental conditions (S-Harm, M-Harm, S-Diss, M-Diss) contained a liquid in the relevant position. The remaining half were distractors, containing no liquids whatsoever (e.g., [dutebi]), such that all stem consonants remained faithful when the suffixes [-li] or [-ɹu] were attached. In Experiment 3, this other half of the training data was replaced with a different set of stems, which likewise stayed faithful when a suffix was added, but which in this case contained a (non-alternating) liquid [l] or [ɹ]. Depending on whether the alternating stems had their alternating liquid in the final or penultimate syllable, the faithful stems contained a liquid in the *other* position. In other words, participants were still given explicit evidence that pairs of co-occurring liquids were subject to harmony at one of the two distances (Short- or Medium-range), but they now also saw evidence that liquids at the other distance *did not harmonize*. As an illustration of the contents of the training phase, **Table 13** provides a breakdown of the number and type of stimuli encountered by the S-Harm-M-Faith group.

  Note that a pattern of this type is fully compatible with the attested *strictly transvocalic* variant of consonant harmony. It is thus expected to be learned as-is, especially given the behaviour of the S-Harm group in Experiment 1 (who inferred non-harmony at Medium-range even without overt evidence). Inclusion of a S-Harm-M-Faith condition is nevertheless important for the interpretability of results. Consider the possibility that the M-Harm-S-Faith group behaves indistinguishably from the Control group. We could not necessarily attribute this failure of learning to the fact that the pattern is more complex than the M-Harm pattern of Experiment 1, since there are additional methodological changes. For example, the fact that all training stems now contain liquids, and that only some of these alternate, might introduce additional processing difficulties that make learning more difficult. It is therefore essential to have a baseline of comparison not only in the Control group, but also the S-Harm-M-Faith group, in order to better interpret the results for the M-Harm-S-Faith group.

7.3. Results and analysis

The response data from the M-Harm-S-Faith and S-Harm-M-Faith groups were subjected to a mixed-effects logistic regression analysis in much the same way as for Experiments 1 and 2, again using the same Control group for a baseline comparison. The structure of the fixed-effects component of the model is identical to that of Experiment 1 (see Section 4.3.1),

**Table 13:** Example of S-Harm-M-Faith training in Experiment 3.

Training triplet	Type and number of items
begeli … begeɹi-ɹu … begeli-li	48 CVCVlV stems with [l]
domelo … domelo-li … domeɹo-ɹu	(Short-range harmony)
mopeɹe … mopeɹe-ɹu … mopele-li	48 CVCVlV stems with [ɹ]
tetoɹi … tetoli-li … tetoɹi-ɹu	(Short-range harmony)
pilede … pilede-li … pilede-ɹu	48 CVlVCV stems with [l]
nelogi … nelogi-ɹu … nelogi-li	(Medium-range faithfulness)
koɹupe … koɹupe-li … koɹupe-ɹu	48 CVlVCV stems with [ɹ]
guɹoto … guɹoto-ɹu … guɹoto-li	(Medium-range faithfulness)

and the random-effects structure (obtained by the same procedure as described there) is minimally different. Full model specifics are provided in the Appendix; here we present only the odds-ratio comparisons of experimental groups against the Control group, shown in **Table 14** (for description of how odds ratios were obtained, see Section 4.3.1). The plot in **Figure 3** shows the performance of each group as well as that of individual participants.

The M-Harm-S-Faith and S-Harm-M-Faith groups both learned liquid harmony at the trigger-target distance where harmony was present in their training data (shaded cells

Table 14: Experiment 3: Odds ratios for choosing harmony at each testing distance, comparing experimental groups to Control group. Shaded cells correspond to contexts where liquid harmony was encountered in training; significance is indicated with boldface.

	Type of test item (trigger-target distance)		
	Short-range (CVCV<u>L</u>V-<u>L</u>V)	**Medium-range** (CV<u>L</u>VCV-<u>L</u>V)	**Long-range** (<u>L</u>VCVCV-<u>L</u>V)
M-Harm-S-Faith versus Control	**2.22** (**p < 0.001**)	**2.50** (**p < 0.001**)	**1.74** (**p = 0.011**)
S-Harm-M-Faith versus Control	**7.42** (**p < 0.001**)	**1.57** (**p = 0.042**)	1.50 (p = 0.062)
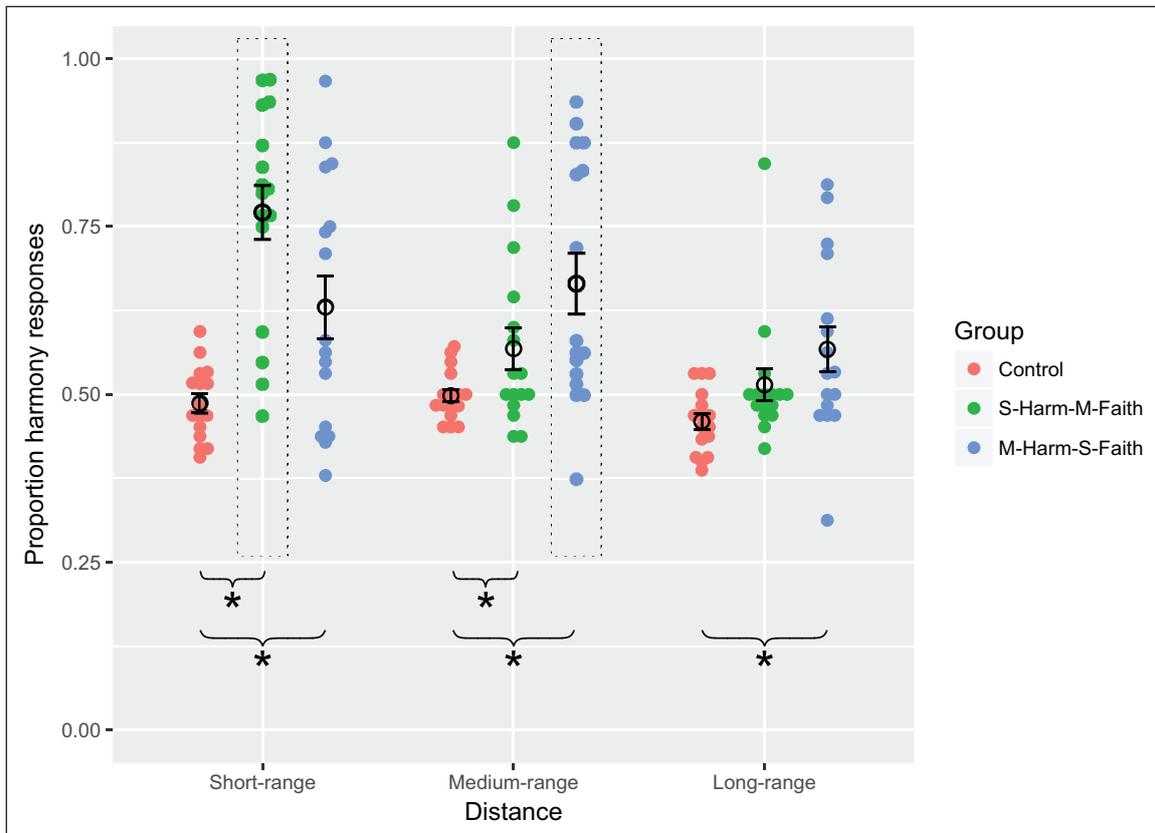
Figure 3: Proportions of harmony responses across conditions (groups) and test-item type (distance) for Experiment 3. Coloured dots show individual participant performance, with group means indicated by an open circle with error bars (showing standard error); significant between-group differences are extracted from the logit mixed model. Performance on contexts where harmony alternations were encountered in training is shown with dashed-line boxes.

in **Table 14**, dashed-line boxes in **Figure 3**). The relevant estimated odds ratios (relative to the Control group) are 2.50 for M-Harm-S-Faith and the considerably higher 7.42 for S-Harm-M-Faith. The small but significant effect for the M-Harm-S-Faith condition is notable, given that this group was exposed to evidence of *non*-harmony in Short-range contexts; evidently this was not enough to stop a number of participants from learning the Medium-range dependency. Visual inspection of **Figure 3** suggests a roughly bimodal distribution for M-Harm-S-Faith participants on Medium-range items, with seven individuals appearing to apply harmony while the remaining nine cluster in the same range as the Control group.

   As a whole, the M-Harm-S-Faith group also chose harmony more often than the Control group in the novel Short-range (OR = 2.22) and Long-range contexts (OR = 1.74). Note that they did so despite explicit evidence in the training data that the liquids in CVCVLV stems *do not* alternate, and that all possible combinations of liquids ([l…l], [[l…r], [r…l], [r…r]]) are permitted in the Short-range CVCVLV-LV contexts. The small but significant effects for both Short- and Long-range test items show that many M-Harm-S-Faith participants overgeneralized harmony from the Medium-range context to the other distances, interpreting the exposure pattern as an *unbounded* dependency despite overt evidence to the contrary. Once again, **Figure 3** depicts an apparently bimodal distribution among M-Harm-S-Faith participants on these types of test-items: Seven individuals tend toward harmony on Short-range items, and four do so on Long-range ones. Not surprisingly, six of the former seven (and all four of the latter) are indeed among the abovementioned seven M-Harm-S-Faith participants who appear to apply harmony in the (familiar) Medium-range context. (See Section 7.4 below for further discussion of individual results).

   A small effect is also present for the S-Harm-M-Faith group on Medium-range items (OR = 1.57). This effect barely reaches statistical significance (p = 0.042) and, judging by **Figure 3**, it stems from a small number of participants who seem to generalize the liquid harmony from Short-range to Medium-range contexts. Nonetheless, this is slightly surprising in light of the fact that this group was exposed to evidence that stem liquids remain faithful in Medium-range contexts. It appears from **Figure 3** that only one S-Harm-M-Faith participant applies harmony in Long-range contexts, and the effect for the group as a whole does not reach significance (OR = 1.50, p = 0.062). The S-Harm-M-Faith results, though peculiar when considering only whether or not an effect reaches statistical significance at the group level, are quite similar to those for the S-Harm group in Experiment 1. In both cases, we see that participants who are exposed to data compatible with strictly transvocalic harmony are very likely to learn that pattern as such, but there is a small tendency (within individuals and, as a result, for their group as a whole) to generalize the dependency to longer distances as well.

   In each of Experiments 1 and 2, we pursued a follow-up analysis examining only those participants who had met a criterion for 'successful learning' of the pattern embedded in their training data. That procedure cannot be straightforwardly extended to the rich-stimulus training conditions of Experiments 3 and 4. Strictly speaking, truly 'successful' learners in the M-Harm-S-Faith and S-Harm-M-Faith conditions would be those showing not only a significant preference for harmony at one distance, but a significant *absence* of any such preference at the other distance as well.

   However, defining and applying such a criterion would not be of interest here. Recall from the predictions outlined in Section 3 that the M-Harm-S-Faith group was expected to fail in learning the pattern embedded in their exposure data (since strictly beyond-transvocalic harmony is both typologically unattested and incompatible with the ABC and

TSL$_2$ hypotheses alike). Such failure could potentially be manifested in several different ways: One participant might detect the Medium-range dependency but erroneously infer an unbounded harmony; another might ignore that dependency and instead generalize the faithful non-alternation from Short-range to other distances; yet another might detect no consistent pattern at all (e.g., responding randomly). It is the first of these patterns that emerges from the aggregated group-level responses, evidently due to a sufficient number of individuals having inferred an unbounded harmony. We emphasize, however, that those participants cannot be considered any more or less 'successful' than those who inferred the *absence* of any harmony dependency (e.g., consistent non-alternation at all distances). Both types of outcome involve accurate learning of the pattern exhibited at one distance level (harmony at Medium-range, or non-alternation at Short-range) combined with overgeneralization of that pattern to the other distance level, in contradiction to the training data. As there was no a priori expectation of one outcome being likelier than the other, a group-level result for M-Harm-S-Faith where harmony was *not* significantly preferred for *any* of the three test-item types would have been equally consistent with the predictions of ABC and TSL$_2$ alike.

For these reasons, we do not single out for special consideration those M-Harm-S-Faith participants who did acquire a pattern involving harmony in Medium-range contexts, nor did we attempt to recruit additional participants who did so for the purpose of further analysis. In the discussion below, we do however highlight certain aspects of the response patterns at the level of individual participants, as these raise some issues of potential interest.

7.4. Discussion of Experiment 3

In all important respects, the results of Experiment 3 agree with those of Experiment 1. Just like the M-Harm group of Experiment 1, the M-Harm-S-Faith group of Experiment 3 tended to learn an *unbounded* harmony pattern, generalizing the Medium-range harmony observed in the training data to Short- and Long-range contexts as well—this despite the fact that half of the training data directly contradicted such an interpretation (the CVCV<u>L</u>V-<u>L</u>V cases). Our results thus indicate a strong learning bias against strictly beyond-transvocalic harmony as a possible type of sound pattern, in line with the ABC and TSL$_2$ hypotheses. These findings, with learners inferring a simpler (or otherwise favoured) pattern in spite of overt evidence for a more complex (or disfavoured) pattern, are analogous to those reported by Lai (2012, 2015) for first-last harmony and by White (2014) for saltatory alternations.

It is worth noting that even though the M-Harm-S-Faith group exhibits a significant preference for harmony at all three distances, the effect sizes are considerably smaller than for the M-Harm group of Experiment 1. The odds ratios relative to the Control group range from 2.40 to 4.67 for the M-Harm group (see **Table 9**), but from 1.74 to 2.50 for the M-Harm-S-Faith group. (Though it is harder to detect, the same difference is evident from the group means for M-Harm and M-Harm-S-Faith in **Figures 1** and **3**, respectively.) This dampening of the overall effect is not surprising, since it was expected that the (presumed) unlearnable status of the strictly beyond-transvocalic harmony pattern embedded in the M-Harm-S-Faith training data might cause a greater number of participants not to learn any consistent pattern at all, or to generalize the Short-range pattern (faithful non-alternation) to Medium-range contexts.

The results for the S-Harm-M-Faith group likewise resemble those of the S-Harm group in Experiment 1, except in that the former exhibit a significant group-level preference for harmony in Medium-range items, suggesting a tendency to generalize from Short-range to

Medium-range contexts (though not to Long-range items). Nevertheless, this effect is very weak and the relevant odds ratios for Medium- and Long-range items are rather similar across the two groups: 1.35 and 1.42, respectively, for S-Harm (see **Table 9**) and 1.57 and 1.50, respectively, for S-Harm-M-Faith (**Table 14**).

It is possible that the composition of the training data in the rich-stimulus conditions of Experiment 3 rendered them more confusing than those in the corresponding Experiment 1 conditions. This could explain the smaller effect sizes in the M-Harm-S-Faith condition as compared to the M-Harm condition of Experiment 1. However, there is reason to be skeptical of such an interpretation. For example, in Short-range items, participants in the (ostensibly confusing) S-Harm-M-Faith condition departed even more decisively from the Control group than did the S-Harm group of Experiment 1: The relevant odds-ratio estimates are 7.42 versus 3.70. This suggests that, if anything, participants' ability to detect the presence of harmony in the Short-range context was aided, rather than undermined, by the addition of training items with liquids in a different position in the stem, and the failure of these liquids to exhibit a harmony alternation. In any case, the close parallel between the results of Experiments 1 and 3 suggests that the rich-stimulus training regime used here is an appropriate design for probing the existence of a learning bias against strictly beyond-transvocalic locality. This makes us more confident in applying the same regime to test whether such a locality pattern is also disfavoured in the case of dissimilation; Experiment 4 investigates this question.

We close this section with some comments on individual results from the M-Harm-S-Faith group in Experiment 3. Full details and plots of these results are provided in the Appendix. In the preceding section, we noted a bimodal distribution whereby only a subset of the sixteen participants appeared to apply harmony while the others' responses clustered in the same range as those of the Control group. Application of the same 95% confidence-limit threshold criterion used in Experiments 1 and 2 (see Section 4.3.2) confirms this: Seven out of the 16 M-Harm-S-Faith participants show evidence of applying harmony in the Medium-range context, and all but one of these seven learners generalize the harmony into Short-range contexts, overriding the counterevidence present in the training data. By the same criterion, only four out of the seven show the same generalization to Long-range contexts; this is similar to Experiments 1 and 2, where participants were likewise reluctant to extend an observed alternation to the word-initial position of L̲VCVCV-L̲V items. However, this relatively clear-cut picture hides a certain amount of idiosyncratic variation at the individual level.

While we emphasize the dangers inherent in attempting to identify and interpret individual differences in a study of this type, the response patterns of two M-Harm-S-Faith participants in particular deserve highlighting. First, one individual (participant 509) appears to have inferred a *strictly transvocalic* harmony, in blatant contradiction to the training data for both Short-range and Medium-range contexts. (The proportions of harmony responses are 0.74, 0.56, and 0.47 on Short-, Medium-, and Long-range items, respectively; only the first of these exceeds the 95% confidence threshold.) One tentative interpretation (keeping in mind the above caveat) might be that this learner did detect that the training data contained a locality-restricted harmony pattern, but reanalyzed the distribution of harmony versus non-harmony contexts in a way that shifted the pattern into the available hypothesis space (the TSL_2 region or, equivalently, the factorial typology of the ABC model).

Secondly, another individual (participant 521) displays a pattern of harmony responses across contexts that does appear to be consistent with a genuine *strictly beyond-transvocalic* harmony. The relevant harmony proportions are 0.43, 0.88, and 0.79 on Short-, Medium-,

and Long-range items, respectively; the latter two are both well above the 95% confidence threshold. As noted above and in Section 1.3, such a pattern is incompatible with both the ABC hypothesis and the TSL_2 hypothesis, and the behaviour of this one participant—if it is anything but a statistical fluke—is therefore surprising. It is of course impossible to conclude much from isolated observations like these, since the experiment was not designed to draw post hoc conclusions from individual participants' deviations from group-level trends. However, we raise them here to shed light on the kinds of underlying individual variation that can at times affect the aggregate group-level results (and their statistical significance). We will return to this issue in the discussion of Experiment 4 in Section 8.4 below, where such individual-level idiosyncracies turn out to pose a bigger problem for interpreting the experimental results.

8. Experiment 4: Locality-restricted dissimilation (rich stimulus)

8.1. Participants, stimuli, and procedure

Thirty-two new participants (21 female, 11 male, mean age 22) were recruited and compensated in the same manner as described for previous experiments, with 16 assigned to each of the M-Diss-S-Faith and S-Diss-M-Faith conditions (described below). We note that an additional 24 participants were subsequently recruited for the M-Diss-S-Faith condition (a follow-up investigation intended to resolve certain ambiguities of interpretation; see Section 8.3 for discussion). Experiment 4 uses the same stimulus set and procedures as in all previous experiments.

8.2. Training conditions

Training items for the two additional experimental conditions were assembled from the stimuli used in the previous experiments, in a manner analogous to Experiment 3 except here the liquid alternation (present in either Short-range or Medium-range items, depending on condition) was one of *dissimilation*, just as in Experiment 2. In each condition, half of the stems had the shape CVCVLV and the other half CVLVCV (96 stems of each type). In the S-Diss-M-Faith condition, dissimilation was seen in CVCVLV-LV items (Short-range context) while the stem liquids stayed faithful in CVLVCV-LV items (Medium-range context). Conversely, in the M-Diss-S-Faith condition the Medium-range items displayed dissimilation while the Short-range items showed faithful non-alternation. **Table 15** provides examples from the training phase of the M-Diss-S-Faith group.

Note that the pattern in **Table 15** complies with the *strictly beyond-transvocalic* variant of dissimilation that is generated within the ABC model (see Section 1.3); it is incompatible with an interpretation that the dependency is *unbounded* (let alone that it is *strictly transvocalic*).

Table 15: Example of M-Diss-S-Faith training in Experiment 4.

Training triplet	Type and number of items
begeli … begeli-ɹu … begeli-li	48 CVCVLV stems with [l]
domelo … domelo-li … domelo-ɹu	(Short-range faithfulness)
mopeɹe … mopeɹe-ɹu … mopeɹe-li	48 CVCVLV stems with [ɹ]
tetoɹi … tetoɹi-li … tetoɹi-ɹu	(Short-range faithfulness)
pilede … piɹede-li … pilede-ɹu	48 CVLVCV stems with [l]
nelogi … nelogi-ɹu … neɹogi-li	(Medium-range dissimilation)
koɹupe … koɹupe-li … kolupe-ɹu	48 CVLVCV stems with [ɹ]
guɹoto … guloto-ɹu … guɹoto-li	(Medium-range dissimilation)

Recall from the predictions laid out in Section 3 that if the ABC model accurately reflects the hypothesis space available to the phonological learner, such a pattern should be learnable by participants in the M-Diss-S-Faith condition. From the perspective of formal language theory, by contrast, all dependencies with strictly beyond-transvocalic locality fall outside of the TSL_2 region no matter whether they involve agreement or disagreement. The TSL_2 hypothesis therefore predicts that strictly beyond-transvocalic dissimilation should be inaccessible to learners, due to its greater computational complexity. The results for the M-Diss-S-Faith group should thus resemble those for the M-Harm-S-Faith group in Experiment 3: Participants should either treat the dissimilation as unbounded, or fail to apply dissimilation at any trigger-target distance. As noted in Section 7.3 above, we have no particular expectation as to the relative prevalence of the former versus the latter outcome among the group. Consequently, an aggregate result for the M-Diss-S-Faith group involving either a consistent preference for dissimilation across all three distance levels or the consistent *absence* of such a preference would be equally consistent with the TSL_2 hypothesis.

8.3. Results and analysis

The response data from the two groups, M-Diss-S-Faith and S-Diss-M-Faith, were analyzed in the same manner as for Experiment 3. Once again, we continue to use the same Control group as a baseline for comparison, and fitted the same kind of mixed-effects logistic regression model over the data from these three groups. As with Experiment 3, we omit the full statistical model here (see the Appendix for details) and instead present the odds ratios extracted from coefficient estimates of the relevant predictors. **Table 16** compares the group-level performance of each experimental group against the Control group for each test-item distance. **Figure 4** plots the proportions of dissimilation responses for each group as well as for individual participants.

Both groups selected test items with liquid disharmony more often than the Control group at the trigger-target distance where that dependency was present in their training data (shaded cells in **Table 16**, dashed-line boxes in **Figure 4**). As expected, the S-Diss-M-Faith group seems to reliably learn the target pattern, as revealed by the large odds ratio of 7.67 on Short-range test items. In the case of the M-Diss-S-Faith group, there is a comparatively weak effect; the odds ratio (relative to the Control group) on Medium-range items is 1.81. Though this effect is statistically significant ($p = 0.009$), visual inspection of **Figure 4** reveals it to be carried entirely by two or three participants who apply Medium-range dissimilation at varying rates (one of them with 100% consistency); other M-Diss-S-Faith participants are clustered near a 0.50 proportion on Medium-range test items (similar to their S-Diss-M-Faith counterparts). Details on individual results from all 16 M-Diss-S-Faith participants are provided for reference in the Appendix.

Table 16: Experiment 4: Odds ratios for choosing disharmony at each testing distance, comparing experimental groups to Control group. Shaded cells correspond to contexts where liquid dissimilation was encountered in training; significance is indicated with boldface.

	Type of test item (trigger-target distance)		
	Short-range (CVCV<u>L</u>V-<u>L</u>V)	Medium-range (CV<u>L</u>VCV-<u>L</u>V)	Long-range (<u>L</u>VCVCV-<u>L</u>V)
M-Diss-S-Faith versus Control	1.04	**1.81**	0.72
	(p = 0.869)	**(p = 0.009)**	(p = 0.118)
S-Diss-M-Faith versus Control	**7.67**	1.33	0.90
	(p < 0.001)	(p = 0.210)	(p = 0.605)
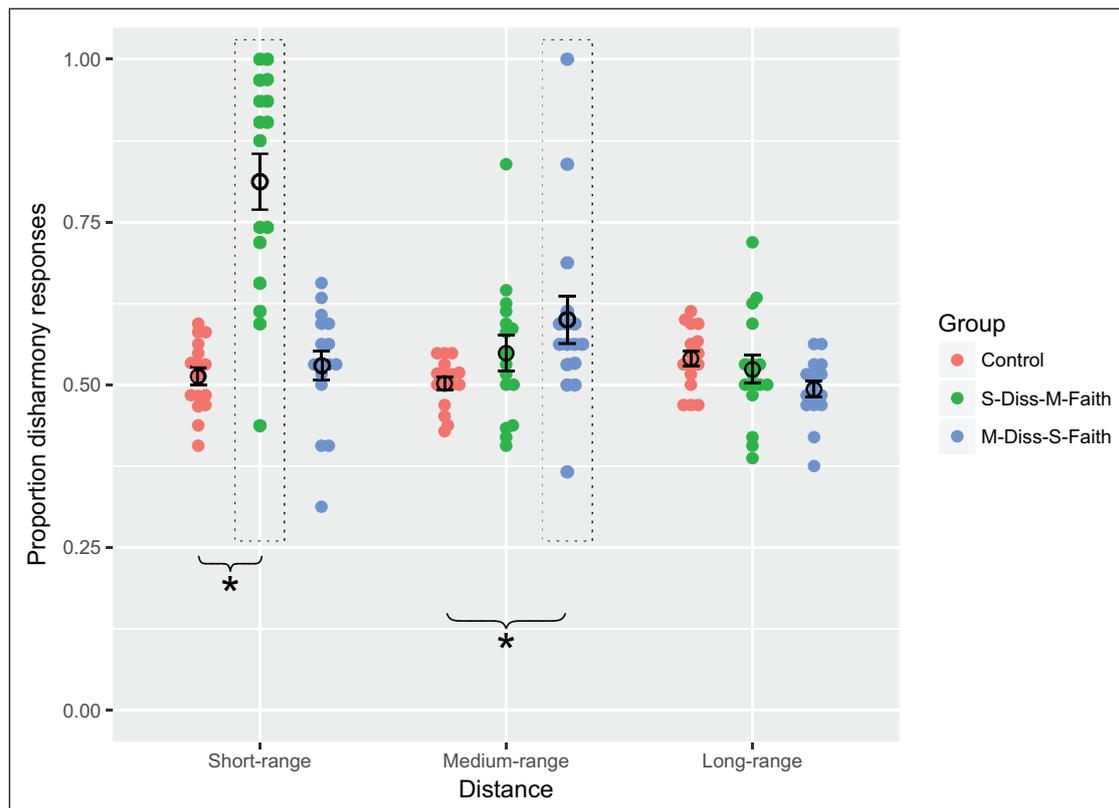
Figure 4: Proportions of disharmony responses across conditions (groups) and test-item type (distance) for Experiment 4. Coloured dots show individual participant performance, with group means indicated by an open circle with error bars (showing standard error); significant between-group differences are extracted from the logit mixed model. Performance on contexts where dissimilation alternations were encountered in training is shown with dashed-line boxes.

As is evident from **Table 16** and **Figure 4**, neither experimental group generalized liquid dissimilation to trigger-target distances other than where it was present in the training data. For the S-Diss-M-Faith group, this pattern of responses is quite robust: As can be seen in **Figure 4**, the vast majority of participants in this group reliably apply dissimilation in Short-range items but show no preference for dissimilation in Medium-range (or Long-range) items. For the M-Diss-S-Faith group, by contrast, the same is not true. Rather, the statistically significant preference for dissimilation in Medium-range items is carried entirely by just two or three individual participants (who, like the remainder of the group, show no analogous preference for dissimilation on Short- or Long-range test items).

As learners' responses to the M-Diss-S-Faith training condition are of central interest, in that they bear on the differing predictions of the ABC and TSL$_2$ hypotheses, we attempted to strengthen the empirical basis for that comparison by recruiting additional participants for the M-Diss-S-Faith condition. The goal was to obtain a minimum of 12 participants who reliably replicated the Medium-range dissimilation featured in the training data (using a 95% confidence-limit threshold criterion). This was analogous to the procedure used in Experiments 1 and 2 for obtaining sufficient numbers of successful learners (see Sections 4.3.2 and 5.3.2); it should nevertheless be kept in mind that in this case, 'successful learner' would not be an accurate characterization (see Section 7.3 for discussion on this point). In what follows, we instead refer to them as Medium-range dissimilators.

However, after running a total of 40 native English speakers in this training condition (including the original 16), the number of participants meeting this criterion had only risen to eight, at which point recruitment was terminated due to practical constraints.

Figure 5 shows the response patterns of the eight Medium-range dissimilators, including the three from the original set of 16 who had sufficiently high rates of Medium-range dissimilation (participants 705, 712, 723). Of these eight, all of whom show a reliable preference for dissimilation on Medium-range items, five surpass the threshold on Short-range items as well (participants 732, 736, 752, 757, 766). In other words, *none* of the Medium-range dissimilators in the additional pool of 24 M-Diss-S-Faith recruits turned out to display the non-generalization to Short-range contexts that was seen for those three who had been in the original group of 16 participants.

8.4. *Discussion of Experiment 4*

The observed behaviour of the S-Diss-M-Faith group was unsurprising and is compatible with both hypotheses under consideration: They did not generalize dissimilation from Short-range contexts to greater distances, and in this respect behaved similarly to their S-Diss counterparts in Experiment 2. The M-Diss-S-Faith results are less clear-cut. As a group, the initial 16 M-Diss-S-Faith participants appear to have acquired a pattern of dissimilation in Medium-range contexts without generalizing it to other distances. This pattern, though non-TSL_2, could be accommodated under the ABC hypothesis, provided there is some independent account of the general reluctance (seen across all the experiments) to generalize alternations into word-initial position (cf. Becker et al., 2012). In other words, at first glance, the aggregate M-Diss-S-Faith results in **Table 16** seem compatible with a characterization that this group learned a *strictly beyond-transvocalic* dissimilation (the effects of which are suppressed in word-initial position), contrary to the TSL_2 hypothesis.

However, as noted in the preceding section, such a conclusion is premature. The weak-but-significant trend in the M-Diss-S-Faith group toward dissimilation on Medium-range items (and the lack of generalization of this dissimilation to Short-range items) is carried entirely by three individuals. Note that these three could truly be called 'successful learners,' in that their responses seem to reflect both aspects of the training data: dissimilation in Medium-range contexts, combined with the absence of dissimilation (i.e., faithful non-alternation) in Short-range contexts. From the point of view of the ABC hypothesis, the small number of individuals who fit this description is quite surprising, especially in the context of other findings in this study. The M-Diss-S-Faith training data were completely consistent with
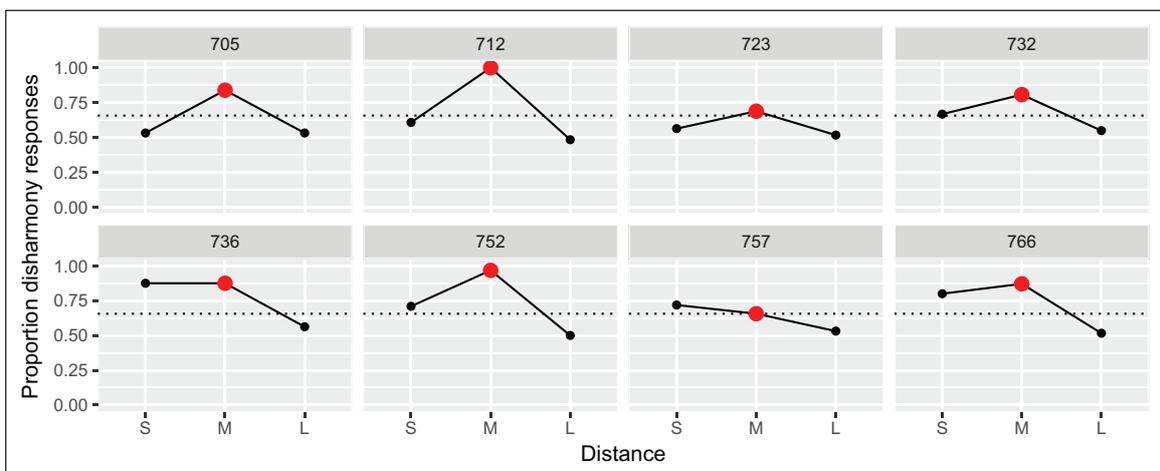
Figure 5: Experiment 4: Individual results for M-Diss–S-Faith participants classified as Medium-range dissimilators, drawing on an expanded pool of 40 participants (see text for discussion). Performance in the context where dissimilation was encountered in training is highlighted in red. The dotted horizontal line indicates the threshold criterion for reliable application of dissimilation.

a possible constraint-ranking configuration in the ABC model (in fact, with several such configurations; see Section 1.3 and the Appendix for details). There is thus no obvious reason why the number of such 'successful learners' (in this sense, i.e., of a *strictly beyond-transvocalic* dissimilation) should differ so sharply from what is seen in the S-Diss-M-Faith condition, or in the S-Harm-M-Faith group of Experiment 3, where the vast majority appear to acquire the *strictly transvocalic* dependency embedded in the training data. A comparison with the M-Harm-S-Faith group of Experiment 3 raises similar questions: That group was exposed to training data that were, by the ABC hypothesis (as well as the TSL$_2$ hypothesis), incompatible with *any* analysis available to the learner, and should thus have caused learners more confusion than the M-Diss-S-Faith training data. Yet we see considerably more M-Harm-S-Faith participants successfully replicating the Medium-range dependency than is the case for the M-Diss-S-Faith group. (A comparison of the M-Harm and M-Diss groups of Experiments 1 and 2 gives no support for such an asymmetry; the effect sizes for learning of the Medium-range dependency were quite similar for both groups; see **Tables 9** and **11**.) We examine the difficulties of capturing these and other asymmetries between training conditions within the ABC model in Section 9.2.1 below.

Secondly, with the effect of interest being driven by such a small number of individuals, the question arises of how reliable that effect is. Recall in this context how the response patterns of individual participants in earlier experiments sometimes deviated from expectations in surprising ways (see Section 7.4). When we attempted to ascertain whether the three M-Diss-S-Faith participants in question were representative of a reliable trend, by expanding the total number of participants in the M-Diss-S-Faith condition from 16 to 40, the number of learners who replicated the Medium-range dissimilation rose from three to eight. All five of these 'new' Medium-range dissimilators *did* generalize the pattern to Short-range contexts. While the number of relevant participants is still too low to draw any decisive conclusions, this suggests that the original group-level results in **Table 16** should be taken with a grain of salt. In the expanded set of data, the majority of Medium-range dissimilators (five out of eight) generalized this dependency to Short-range items, in contradiction to the evidence provided in the training data, just as did the majority of learners of the Medium-range harmony (six out of seven) in the M-Harm-S-Faith condition of Experiment 3 (see Section 7.3). Once again, this apparent similarity between the M-Diss-S-Faith and M-Harm-S-Faith conditions is unexpected under the ABC model, given that the training data for the former, but not the latter, were fully consistent with the hypothesis space that is presumed to be available to the learner.

In sum, we do not believe that the results of Experiment 4 should be interpreted as support for the ABC hypothesis over the TSL$_2$ hypothesis. On the contrary, data from the expanded set of M-Diss-S-Faith participants suggests that Medium-range dissimilators do in fact tend to overgeneralize this into transvocalic contexts (i.e., infer an unbounded rather than strictly beyond-transvocalic dependency). This behaviour is comparable to the results of the M-Harm-S-Faith group of Experiment 3 and is consistent with the predictions of the TSL$_2$ hypothesis.

9. General discussion

The purpose of this series of experiments was to examine how humans learn and generalize various types of long-distance phonotactic dependencies. We find strong evidence of learning biases (inductive biases) that seem to guide learners' hypothesis formation about the dependency patterns they are exposed to. Here we summarize the biases that receive support from the experimental results (Section 9.1), before turning to their implications for formal theories of long-distance phonological dependencies (Section 9.2). Finally, we highlight some remaining issues for further research that arise from our study (Section 9.3).

9.1. Evidence for learning biases

The results of Experiments 1 and 2 point to a hierarchy in terms of which locality parameters are preferred by learners, both for long-distance consonant assimilation and dissimilation: No dependency > Strictly transvocalic dependency > Unbounded dependency > Strictly beyond-transvocalic dependency. The evidence suggests that this is the order in which a typical learner will consider each of these hypotheses; in the face of ambiguous evidence, learners tend to interpret the data conservatively on the above scale. For example, a training phase which contains no stems with liquids whatsoever is, in principle, compatible with any type of dependency between liquids. The Control group, who were exposed to just such input, showed no preference for harmony or dissimilation at any of the test-item distances. Likewise, the training phases for the S-Harm and S-Diss conditions (in Experiments 1 and 2, respectively) contained no evidence about liquid-liquid co-occurrences at Medium- or Long-range distances, and were thus consistent with either a strictly transvocalic or an unbounded dependency. However, participants in these conditions were reluctant to extend the inferred pattern to distances beyond the …CVC… window, preferring instead to interpret the dependency as strictly transvocalic. Finally, the M-Harm and M-Diss groups (Experiments 1 and 2) were exposed to a Medium-range dependency only, with no evidence bearing on Short- or Long-range distances. While this input was consistent with an interpretation of the dependency as either strictly beyond-transvocalic or unbounded, participants preferred the latter, generalizing to the Short-range and also (though somewhat less reliably) the Long-range distances.

The results of Experiments 3 and 4 suggest that patterns with strictly beyond-transvocalic locality, which appear to be cross-linguistically unattested (for the questionable case of Sundanese dissimilation, see Section 1.4) are indeed extremely difficult for phonological learners to access. Only a small number of participants displayed individual results that seem consistent with such a pattern (one M-Harm-S-Faith participant, three M-Diss-S-Faith participants). However, most had difficulty detecting any Medium-range dependency at all (in contrast to the M-Harm and M-Diss conditions of Experiments 1 and 2, where participants had no trouble picking up on interactions at that distance). Of those who did, the majority over-generalized to Short-range contexts as well, thus deviating from the pattern shown to them in training. This suggests that strictly beyond-transvocalic locality patterns are situated at the outer limit of, if not beyond, the learner's available hypothesis space and that any viable theory of long-distance dependencies should be reflective of this bias.

9.2. Theoretical implications

9.2.1. Agreement by Correspondence

To address whether the learning biases outlined above can be accommodated within the Agreement by Correspondence model (ABC; see Sections 1.2.2 and 1.3), we first note that there are different types of bias that play a role in formal constraint-based models of phonology. The first type of bias to consider is one that defines the boundary of the learner's hypothesis space; in Optimality Theory this is provided by the universal *constraint set*. Since the learner's goal is to discover the appropriate ranking of a finite number of constraints, patterns that cannot be generated within the factorial typology of that constraint set lie outside of the available hypothesis space.

Given a basic set of relevant constraints from the ABC model, the typology of liquid-liquid dependencies that can be generated, and should thus in principle be learnable, is tabulated in (10) (cf. **Table 2**). We used the OT-Help software (Staubs et al., 2010) to compute the full typology, due to the large number of possible ranking permutations of the eight constraints that were included (IO-IDENT[lateral], CORR-LIQ, CORR-LIQ$_{CVC}$,

CORR-LIQ[αlateral], CORR-LIQ[αlateral]$_{CVC}$, PROXIMITY, CC-IDENT[lateral], CC-NOCORR; see Sections 1.2.2 and 1.3 above). Details are provided in the Appendix, to which the reader is referred for precise definitions of the constraints used and descriptions of the crucial ranking relations that define each of the patterns in (10); the OT-Help source files are also provided as supplementary materials available online.

(10)   Factorial typology (hypothesis space) generated by the ABC constraint set
a.   No dependency
b.   Unbounded harmony
c.   Unbounded dissimilation
d.   Strictly transvocalic harmony
e.   Strictly transvocalic dissimilation
f.   Strictly beyond-transvocalic dissimilation
g.   Transvocalic harmony with beyond-transvocalic dissimilation

In some respects, the typology in (10) fits with both the cross-linguistic evidence and our experimental results.[8] For instance, it successfully omits strictly beyond-transvocalic harmony, which is unattested and which was not reliably learned in Experiment 1 (M-Harm condition) or Experiment 3 (M-Harm-S-Faith condition). As already discussed in Section 1.3, however, the typology generated by ABC does include strictly beyond-transvocalic *dissimilation* (10f), for which the support is at best weak, both cross-linguistically and in our experimental results (Experiment 2, M-Diss condition; Experiment 4, M-Diss-S-Faith condition). Finally, it is worth noting that a strictly *transvocalic* dissimilation pattern (10e) can only be derived in the ABC model under certain assumptions about its constraint set. Specifically, that set must include special versions of the CORR-X constraints (which require correspondence among members of some natural class X) that refer only to transvocalic …CVC… consonant pairs. The inclusion of such locality-sensitive CORR-X constraints was originally proposed to account for transvocalic harmony (Hansson, 2010) and later adopted for transvocalic dissimilation (Bennett, 2015b). As strictly transvocalic dissimilation is robustly attested (Bennett, 2015b), and was reliably learned in our experiments (S-Diss and S-Diss-M-Faith conditions of Experiments 2 and 4, respectively), such constraints are essential if ABC is to be an empirically viable model. As we noted in Section 1.3, the inclusion or exclusion of such constraints has no bearing on the ABC model's prediction that strictly *beyond-transvocalic* locality is learnable for dissimilation (10f) but not for harmony (for more details on this point, see the Appendix).

  While the factorial typology reflects a categorical bias that determines which patterns are learnable and which are not, the *initial ranking* (or weighting) of constraints can also shape the learning process, as it dictates the learner's starting point and trajectory through the hypothesis space. In the absence of any evidence motivating a re-ranking of two or more constraints, the initial grammar will generate a default pattern. The fact that the Control group showed no preference for harmony or dissimilation at any of the three testing distances (and largely favoured whichever item was faithful to the bare verb stem) suggests that some constraint that militates against alternations is undominated in the initial ranking. An output-output faithfulness constraint OO-IDENT[lateral], which penalizes divergence from a related base form (e.g., past tense [pido̠ɹe-ɹu], given present tense [pido̠le]), can represent this anti-alternation bias.[9]

In order for a learner to transition to a grammar that generates liquid harmony or dissimilation alternations, OO-IDENT[lateral] must be demoted below those Markedness constraints that trigger harmony or dissimilation. For liquid harmony, these are CORR-LIQ (and/or its locality-restricted counterpart CORR-LIQ$_{CVC}$), which require co-occurring liquids to stand in correspondence, and CC-IDENT[lateral], which requires correspondent segments (i.e., liquids) to agree in [±lateral]. For liquid dissimilation, the relevant constraints are CORR-LIQ[αlateral] (and/or CORR-LIQ[αlateral]$_{CVC}$), which demands correspondence in co-occurring *identical* liquids ([l…l], [ɹ…ɹ]), as well as either PROXIMITY/CC-SYLLADJ (which prohibits correspondence across one or more intervening syllables) or CC-NOCORR (shorthand for an effectively across-the-board ban on correspondence; see Section 1.3) or both. (For details on constraint definitions and crucial ranking relations, see the Appendix).

The apparent strictly-transvocalic > unbounded locality preference, which emerged for both harmony (Experiment 1) and dissimilation (Experiment 2), can similarly be approximated in the ABC model. Strictly transvocalic dissimilation and harmony alike can be derived under a general ranking schema CORR-X$_{CVC}$ ≫ OO-IDENT[F] ≫ CORR-X, where X defines the class of interacting segment pairs, and [F] is the assimilating or dissimilating feature. (The difference between harmony and dissimilation under such a schema falls to the relative ranking among CC-IDENT[F], CC-NOCORR, and the relevant faithfulness constraints IO-IDENT[F], OO-IDENT[F], as well as how narrowly the class X is defined.) The observed preference for strictly-transvocalic over unbounded locality could thus be captured by positing that the locality-restricted (CVC) version of any CORR-X constraint outranks its unrestricted counterpart in the learner's initial state.[10]

Capturing the dispreference for strictly *beyond*-transvocalic locality relative to all other types (including unbounded locality) is more problematic. As described in Section 1.3, strictly beyond-transvocalic dissimilation can arise when the constraint PROXIMITY/CC-SYLLADJ is undominated; we might therefore propose that this constraint is ranked/weighted quite low in the initial state—or even that it be eliminated from the constraint set altogether. However, strictly beyond-transvocalic dissimilation can still be generated without any involvement of this constraint (whereas strictly beyond-transvocalic harmony cannot). Furthermore, it too arises from a ranking schema of the type CORR-X$_{CVC}$ ≫ C ≫ CORR-X (where C is some other constraint), just like strictly transvocalic dissimilation does. When C is a correspondence-penalizing constraint like CC-NOCORR, the result is a dissimilation that is confined to beyond-transvocalic contexts: Transvocalic pairs are forced to correspond (thereby escaping dissimilation) while more distant pairs are forced *not* to correspond (and hence must dissimilate, to evade the need for correspondence). Strictly transvocalic dissimilation and unbounded dissimilation both require CC-NOCORR to outrank both versions of the relevant CORR-X constraint. It will not do to posit that this latter ranking holds in the initial state, in the hope of disfavouring strictly *beyond*-transvocalic dissimilation relative to these other locality types, as this makes highly undesirable predictions. In particular, any *harmony* dependency—be it transvocalic or unbounded—requires CC-NOCORR to be ranked *below* all CORR-X constraints. We would thus expect all harmony patterns to pose an even bigger challenge to learners than the least

---

White, 2014). Error-driven constraint-ranking algorithms typically rely on a Markedness ≫ IO-Faithfulness bias to ensure a restrictive final grammar, especially for phonotactic generalizations not supported by alternations (Hayes, 2004; Jesney & Tessier, 2011; Prince & Tesar, 2004; Smolensky, 1996). For this reason it is less feasible to attribute the anti-alternation bias to high ranking of an IO-Faithfulness constraint like IO-IDENT[lateral].

[10] With weighted rather than ranked constraints, this effect is achieved for free, due to the inherent subset–superset relation between the two versions of each CORR-X constraint: Anything that violates the CVC version also violates the general version, but not vice versa. As a result, the bias for strictly transvocalic over unbounded locality holds regardless of the relative weighting of the two constraint versions in the initial state.

accessible of all dissimilation patterns, namely strictly beyond-transvocalic dissimilation. This is in no way consistent with our findings for Experiments 1 and 3 as compared to Experiment 4.

In conclusion, the ABC model predicts a sharp divide between dissimilation and harmony in terms of the learnability of strictly beyond-transvocalic dependencies: While such a locality pattern lies within the learner's hypothesis space in the case of dissimilation, it is predicted to be categorically unlearnable for harmony. Our results from Experiments 3 and 4 (and, less conclusively, Experiments 1 and 2) do not support such a predicted mismatch between these two types of long-distance dependencies. Furthermore, it does not seem possible to make the required ranking configuration for strictly beyond-transvocalic locality (for dissimilation) relatively inaccessible, by formulating a learning bias in the form of an initial constraint ranking (or weighting) among the constraints of the ABC model. Any such attempt leads to even more problematic predictions regarding the learners' preferences in other respects, which are similarly inconsistent with our findings (e.g., the general preference for strictly transvocalic over unbounded locality, for dissimilation and harmony alike).

9.2.2. Formal language theory

The range of learning biases described above in Section 9.1 are more easily explained when the target patterns are defined in terms of TSL_2 grammars. Recall from Section 1.2.3 that the formal grammar of a TSL_2 language (stringset) is a two-tuple $G = \langle T, R \rangle$, defined over some alphabet Σ (a segment inventory), where the tier T is some subset of Σ over which adjacency is assessed, and R is the set of bigrams that are prohibited on that tier. Distinctions among the types of patterns that participants evidently learned can easily be characterized by varying different aspects of the formal grammar; this is summarized in **Table 17**. As noted in Sections 1.2.3 and 1.3, the difference between liquid agreement (harmony) and disagreement (dissimilation) lies in the set of restricted bigrams R, whereas the distinction between unbounded and strictly transvocalic dependencies arises from whether the tier T is limited to those segments that occur in R or includes other consonants as well.[11]

On the other hand, the strictly beyond-transvocalic locality pattern indicated by the exposure data for the M-Harm-S-Faith and M-Diss-S-Faith conditions of Experiments 3 and 4, respectively, cannot be characterized in TSL_2 terms. Recall from Section 1.2.3 that such a dependency, which holds across *at least one* consonant, is not TSL_k for any value of k, no matter how T and R are construed. Rather, a formal characterization of strictly beyond-transvocalic locality requires at least a Locally Testable (LT) grammar. For example, the strictly beyond-transvocalic liquid harmony pattern the M-Harm-S-Faith

Table 17: TSL_2 grammars for long-distance dependencies between liquids.

Locality	Type	Σ	T	R
unbounded	agreement	{l, r, C, V}	{l, r}	{*lr, *rl}
	disagreement	{l, r, C, V}	{l, r}	{*ll, *rr}
transvocalic	agreement	{l, r, C, V}	{l, r, C}	{*lr, *rl}
	disagreement	{l, r, C, V}	{l, r, C}	{*ll, *rr}

---

[11] In this case, the difference is whether T includes *all* or *none* of the consonants that do not figure in R (here, non-liquids). If T includes only a subset of those other consonants, the pattern generated is a dependency that is blocked by members of that subset but is otherwise unbounded (McMullin, 2016). Such blocking effects are attested both for consonant harmony (Hansson, 2010) and dissimilation (Bennett, 2015b).

group encountered can be described by a (tier-based) LT_2 grammar, in which words must not simultaneously contain, on the tier of consonants (and boundary symbols, $T = \{l, r, C, \#\}$), a member of the bigram set $\{\#l, Cl\}$ as well as a member of the bigram set $\{\#r, Cr\}$. Given the markedly greater computational complexity of strictly beyond-transvocalic locality, the general failure of participants in the relevant conditions of Experiments 3 and 4 to learn such dependency patterns is unsurprising.[12] Our results are in line with those of Lai (2015), who found that participants were similarly unable to learn a pattern of 'first-last' harmony (which is likewise LT_2 but not TSL_2; see Section 1.2.3). These findings suggest that the boundary of the TSL_2 region constitutes an important hurdle for human sound-pattern learning.

Turning now to the observed preference relations among locality patterns, these too fit well with a characterization in terms of language-theoretic complexity classes. First, since the training items for the Control group contained no stems with liquids, there was no reason for them to prefer a grammar that generates either liquid harmony or liquid dissimilation.[13] Secondly, the learners' apparent preference for strictly transvocalic over unbounded locality (when faced with ambiguous evidence) translates into a preference for hypothesizing a tier containing all consonants over one that contains only a subset of consonants (namely liquids), other things being equal. Interestingly, the formal learning algorithm presented by Jardine and McMullin (2017), which is proven to be successful for the entire TSL_k class (i.e., for any given k), begins with the hypothesis that $T = \Sigma$ (see also Jardine & Heinz, 2016). It then removes segments from the tier one-by-one when possible, on the basis of overt evidence that they do not themselves figure in R and that co-occurrence restrictions hold across them (i.e., that they are transparent, not blockers). In this way, the trajectory of the algorithm parallels how the S-Harm (Exp. 1) and S-Diss (Exp. 2) participants were biased toward interpreting otherwise ambiguous evidence as indicative of a strictly transvocalic pattern, which is specified with more tier-based segments than its unbounded equivalent.[14]

In sum, we argue that the Tier-based Strictly 2-Local class of formal languages reflects the range of attested long-distance dependencies between consonants and provides a transparent account of our experimental findings. In particular, the observed parallels

---

[12] In fact, the strictly beyond-transvocalic dissimilation pattern that the M-Diss-S-Faith group in Experiment 4 were exposed to is even more complex than its harmony counterpart (M-Harm-S-Faith in Experiment 3). The former cannot be captured by a (tier-based) LT_2 grammar, because it requires *counting* of occurrences of members of the relevant bigram sets. For example, in order to rule out strings of the structure *…(CV)rVCV…rV… while permitting …(CV)rVrV…, it is necessary to prohibit words from containing *two or more instances* of bigrams from the set $\{\#r, Cr\}$ on the consonant tier (and similarly for $\{\#l, Cl\}$, in the case of [l…l] dissimilation). Strictly beyond-transvocalic *dissimilation* thus belongs to the even more complex Locally Threshold-Testable (LTT) class. It could therefore be said that the formal language-theoretic perspective predicts a learnability mismatch between harmony and dissimilation under strictly beyond-transvocalic locality, just like ABC does, but in the exact opposite direction: Beyond-transvocalic dissimilation should be *harder* to learn than harmony, not vice versa. Since neither type of strictly beyond-transvocalic dependency was learned successfully in our experiments, our results do not bear on this inherent difference in complexity.

[13] In principle, a learner might infer from the mere absence of any forms with co-occurring liquids that all such co-occurrences were banned. This amounts to positing that $R = \{*lr, *rl, *ll, *rr\}$ on a tier $T = \{l, r\}$. However, in each testing trial, both response options violated this phonotactic restriction, and hence no detectable pattern of responses should follow from a learner having formed such a hypothesis.

[14] Strictly speaking, the algorithm proposed by Jardine and McMullin (2017) would not actually learn the target tier if it was given the exact same training data provided to participants. The reason for this is that TSL grammars may specify exactly one tier, whereas characterizing the overall phonotactics of the training stimuli requires *multiple* tiers. Specifically, since all stimuli have CV syllable structure, the algorithm would simply infer a tier consisting of all segments (i.e., the largest possible tier) with each of the bigram restrictions $\{CC, CL, LC, LL, VV\}$. For further discussion of accommodating and learning phonotactic restrictions on multiple tiers within a single formal grammar, see McMullin (2016), as well as McMullin, Aksënova, and De Santo (2019).

between the learning of harmony and dissimilation (which are problematic for the ABC hypothesis; see Section 9.2.1) are entirely expected, given that these two types of dependencies reflect a minor distinction among formal grammars of equivalent complexity. The drastic difficulty exhibited by learners in the relevant conditions of Experiments 3 and 4 in detecting and generalizing a strictly beyond-transvocalic dependency pattern, for harmony and dissimilation alike, is compatible with the TSL_2 hypothesis whereas it is unexpected under ABC. Moreover, existing TSL learning algorithms (in particular that of Jardine & McMullin, 2017) not only induce the contents of the tier itself, but do so in a way that mirrors our experimental findings. Corresponding learnability results do not exist for constraint-based learning algorithms under the ABC model (neither with the set of all conceivable ABC constraints given in advance, nor with induction of appropriate constraints from observed data). This is a non-trivial issue, given the hidden-structure character of the abstract correspondence relations on which ABC relies.

9.3. Further issues

The discussion of language-theoretic classes and complexity relations in preceding sections has been framed exclusively in terms of phonotactics and phonotactic learning. This may appear to be at odds with the fact that the experiment itself involved morpho-phonological *alternations* between forms. Admittedly, there is not yet a deep understanding of the formal and learnability properties of subregular functions that can model string-to-string mappings. However, recent work has defined a number of relevant classes and identified some of their formal properties (e.g., Chandlee, 2014; Chandlee, Eyraud, & Heinz, 2015; Mohri, 1997), including a number of promising results regarding learnability (Chandlee, Eyraud, & Heinz, 2014; Chandlee & Jardine, 2014). Of particular relevance is the finding that subsequential functions (Mohri, 1997) characterize many attested patterns of long-distance consonant dissimilation (Payne, 2014) and consonant harmony (Luo, 2017), while excluding a number of complex pathologies. Further work will determine whether or not the class of TSL functions defined by Chandlee, Heinz, Jardine, and McMullin (2017), which are more restrictive than the subsequential functions, can be learned with a technique similar to that employed by Jardine and McMullin (2017).

As noted in Section 1.1, we have sidestepped the question of whether the dependencies we have described as strictly 'transvocalic' (as well as strictly beyond-'transvocalic') are based on consonant-tier adjacency or syllable adjacency. Our formal-language-theoretic characterizations of such patterns have assumed the former; they are defined over strings of segments, with no reference to syllable constituency.[15]

For the ABC model, however, things are more complex: The most elaborated and formally explicit version of the model, that of Bennett (2015b), contains constraints referencing each of these two locality criteria. On the one hand, the CC-Limiter constraint CC-SYLLADJ (equivalent to PROXIMITY in Rose & Walker, 2004) bans correspondence between segments that are separated by an intervening syllable, i.e., are not *syllable-adjacent.* On the other hand, special CVC versions of the correspondence-inducing CORR-X constraints are posited (following Hansson, 2010), which apply only to consonant pairs that are *consonant-tier-adjacent.* As discussed in Section 9.2.1, each of these constraint types

---

[15] If future empirical evidence were to point in favour of the syllable-based alternative, words could instead be defined as strings of syllables. For a TSL_2 characterization of, e.g., liquid harmony in these terms, the alphabet Σ would not include any syllables that contain both [l] and [r], such as {lVr, rVl, ClVr, lVrC}, and the restrictions R would consist of all syllable bigrams where one syllable contains [l] and the other [r]. The distinction in locality would again result from differences in the tier T. For an unbounded harmony, T would contain exactly the set of syllables that contain either [l] or [r], while for a strictly transvocalic harmony, all syllables would be on the tier (such that a word like *[lVC.CVr] would be banned, but [lVC.CV.rV] would not).

can give rise to 'transvocalic' harmony, as well as to beyond-'transvocalic' dissimilation, whereas only the CORR-X$_{CVC}$ constraint type can generate 'transvocalic' dissimilation. A true characterization of the factorial typology of the ABC model as per Bennett (2015b) is thus more nuanced. While harmony can be either strictly consonant-tier-adjacent, strictly syllable-adjacent, or unbounded, dissimilation can be strictly consonant-tier-adjacent, strictly beyond-consonant-tier-adjacent, strictly beyond-syllable-adjacent, or unbounded. Notably, while strictly *syllable*-adjacent locality is available for harmony, it is predicted to be impossible for dissimilation patterns. It is not clear whether this particular mismatch prediction for harmony versus dissimilation is intentional on the part of Bennett (2015b); we are not aware of any cross-linguistic basis for it.

10. Conclusions

This paper investigated the role of inductive biases in the learning of long-distance interactions between consonants and the extent to which these biases reflect the cross-linguistic typology of locality relations for such patterns. In a series of artificial grammar learning experiments focused on liquid harmony (Experiments 1 and 3) and liquid dissimilation (Experiments 2 and 4), we have both replicated and expanded upon previous findings (Finley, 2011, 2012; McMullin & Hansson, 2014), allowing us to better evaluate current theoretical models of long-distance dependencies in segmental phonology.

In line with the cross-linguistic typology, participants were able to learn both strictly transvocalic and unbounded versions of consonant harmony (Experiment 1) and dissimilation (Experiment 2). The logically possible, but formally more complex, strictly beyond-transvocalic patterns of consonant harmony (Experiment 3) and dissimilation (Experiment 4) were not reliably acquired. Even though unambiguous evidence for such patterns was provided, participants tended instead to infer a simpler version (unbounded locality) or failed to learn any pattern at all. In sum, the experiments revealed no differences between assimilatory and dissimilatory dependencies with regard to learning of different locality patterns.

These findings are consistent with the hypothesis that the Tier-based Strictly 2-Local class of formal languages (TSL$_2$; Heinz et al., 2011) delimits the range of learnable phonotactic patterns (McMullin, 2016; McMullin & Hansson, 2015; cf. Lai, 2015). Assimilatory and dissimilatory co-occurrence restrictions within the TSL$_2$ region are of equivalent complexity, and should thus not differ with respect to learnability, whereas strictly beyond-transvocalic dependencies of either type fall outside of the TSL$_2$ region and hence pose a greater (perhaps insurmountable) challenge to the learner. Moreover, a general preference for strictly transvocalic over unbounded locality, which was observed for both harmony and dissimilation, is predicted by existing algorithms that learn TSL grammars (Jardine & Heinz, 2016; Jardine & McMullin, 2017). By contrast, the results cannot be said to support the predictions of the constraint-based Agreement by Correspondence model of long-distance segmental interactions (ABC; Bennett, 2015b; Hansson, 2010; Rose & Walker, 2004), in which strictly beyond-transvocalic locality should be learnable for dissimilation but not harmony. We conclude that, with respect to long-distance consonant interactions, the TSL$_2$ region of the subregular hierarchy better approximates the phonological learner's hypothesis space than the factorial typology defined by the ABC model.

Additional Files

The additional files for this article can be found as follows:

- **File 1.** An appendix (PDF file) providing further details on many aspects of this article. Section A1 outlines additional aspects of the experimental procedures. Section A2 provides a full list of stimuli used in Experiments 1–4. Section A3 describes the

statistical methods and presents each individual model in full. Plots depicting individual performance for participants in Experiments 3 and 4 are shown in Section A4. Finally, Section A5 provides additional details about how the factorial typology of ABC constraints was generated. DOI: https://doi.org/10.5334/labphon.150.s1

- **File 2.** A ZIP archive containing the source files (`ABCDissHarmFull.txt`, `ABCDissHarmNoCVC.txt`, and `ABCDissHarmNoProxim.txt`) that were used in generating the factorial typologies of ABC constraints as described in Section A5 of the appendix (File 1). DOI: https://doi.org/10.5334/labphon.150.s2

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Baer-Henney, D., Kügler, F., & van de Vijver, R. (2015). The interaction of language–specific and universal factors during the acquisition of morphophonemic alternations with exceptions. *Cognitive Science*, *39*(7), 1537–1569. DOI: https://doi.org/10.1111/cogs.12209

Baer-Henney, D., & van de Vijver, R. (2012). On the role of substance, locality and amount of exposure in the acquisition of morphophonemic alternations. *Laboratory Phonology*, *3*(2), 221–249. DOI: https://doi.org/10.1515/lp-2012-0013

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. *ArXiv e-prints*. Retrieved from https://arxiv.org/abs/1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Becker, M., Nevins, A., & Levine, J. (2012). Asymmetries in generalizing alternations to and from initial syllables. *Language*, *88*(2), 231–268. DOI: https://doi.org/10.1353/lan.2012.0049

Beguš, G. (2018). *Unnatural phonology: A synchrony-diachrony interface approach* (Doctoral dissertation). Harvard University.

Bennett, W. G. (2015a). Assimilation, dissimilation, and surface correspondence in Sundanese. *Natural Language and Linguistic Theory*, *33*(2), 371–415. DOI: https://doi.org/10.1007/s11049-014-9268-2

Bennett, W. G. (2015b). *The phonology of consonants: Harmony, dissimilation and correspondence*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139683586

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511486357

Carpenter, A. C. (2010). A naturalness bias in learning stress. *Phonology*, *27*(3), 345–392. DOI: https://doi.org/10.1017/S0952675710000199

Chandlee, J. (2014). *Strictly Local phonological processes* (Doctoral dissertation). University of Delaware.

Chandlee, J., Eyraud, R., & Heinz, J. (2014). Learning Strictly Local subsequential functions. *Transactions of the Association for Computational Linguistics, 2*, 491–503. DOI: https://doi.org/10.1162/tacl_a_00198

Chandlee, J., Eyraud, R., & Heinz, J. (2015, July). Output Strictly Local functions. In *Proceedings of the 14th Meeting on the Mathematics of Language (MoL 2015), 2*, 112–125. Chicago, IL: Association for Computational Linguistics. DOI: https://doi.org/10.3115/v1/W15-2310

Chandlee, J., Heinz, J., Jardine, A., & McMullin, K. (2017). Modeling long-distance alternations with Tier-based Strictly Local functions. *Paper presented at the 91st Annual Meeting of the Linguistic Society of America*. Austin, TX.

Chandlee, J., & Jardine, A. (2014). Learning phonological mappings by learning strictly local functions. In J. Kingston, C. Moore-Cantwell, J. Pater, & R. Staubs (Eds.), *Proceedings of the 2013 Meeting on Phonology*. Washington, DC: Linguistic Society of America. DOI: https://doi.org/10.3765/amp.v1i1.13

Coetzee, A. W. (2009). Learning lexical indexation. *Phonology, 26*, 109–145. DOI: https://doi.org/10.1017/S0952675709001730

Cohn, A. C. (1992). The consequences of dissimilation in Sundanese. *Phonology, 9*(2), 199–220. DOI: https://doi.org/10.1017/S0952675700001585

Cristia, A., Mielke, J., Daland, R., & Peperkamp, S. (2013). Similarity in the generalization of implicitly learned sound patterns. *Laboratory Phonology, 4*(2), 259–285. DOI: https://doi.org/10.1515/lp-2013-0010

Culbertson, J. (2012). Typological universals as reflections of biased learning: Evidence from artificial language learning. *Language and Linguistics Compass, 6*(5), 310–329. DOI: https://doi.org/10.1002/lnc3.338

Do, Y. A. (2013). *Biased learning of phonological alternations* (Doctoral dissertation). Massachusetts Institute of Technology.

Dolbey, A., & Hansson, G. Ó. (1999). The source of naturalness in synchronic phonology. In S. J. Billings, J. P. Boyle, & A. M. Griffith (Eds.), *Papers from the 35th meeting of the Chicago Linguistic Society, 1*, 59–69. Chicago, IL: Chicago Linguistic Society.

Fallon, P. D. (1993). Liquid dissimilation in Georgian. In A. Kathol, & M. Bernstein (Eds.), *Proceedings of the 10th Eastern States Conference on Linguistics (ESCOL 10)* (pp. 105–116). Ithaca, NY: DMLL Publications. DOI: https://doi.org/10.1075/avt.10.12ros

Finley, S. (2011). The privileged status of locality in consonant harmony. *Journal of Memory and Language, 65*, 74–83. DOI: https://doi.org/10.1016/j.jml.2011.02.006

Finley, S. (2012). Testing the limits of long-distance learning: Learning beyond a three-segment window. *Cognitive Science, 36*, 740–756. DOI: https://doi.org/10.1111/j.1551-6709.2011.01227.x

Finley, S. (2015). Learning non-adjacent dependencies in phonology: Transparent vowels in vowel harmony. *Language, 91*(1), 48–72. DOI: https://doi.org/10.1353/lan.2015.0010

Finley, S. (2017). Locality and harmony: Perspectives from artificial grammar learning. *Language and Linguistics Compass, 11*(1). DOI: https://doi.org/10.1111/lnc3.12233

Finley, S., & Badecker, W. (2009). Artificial language learning and feature-based generalization. *Journal of Memory and Language, 61*(3), 423–437. DOI: https://doi.org/10.1016/j.jml.2009.05.002

Gallagher, G. (2013). Learning the identity effect as an artificial language: Bias and generalisation. *Phonology, 30*(2), 253–295. DOI: https://doi.org/10.1017/S0952675713000134

Garrett, A., & Johnson, K. (2013). Phonetic bias in sound change. In A. C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 51–97). Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199573745.003.0003

Glewwe, E. (2019). Substantive bias in phonotactic learning: Positional extension of an obstruent voicing contrast. In *Papers from the 53rd regional meeting of the Chicago Linguistic Society* (pp. 103–116). Chicago, IL: Chicago Linguistic Society.

Glewwe, E., Zymet, J., Adams, J., Jacobson, R., Yates, A., Zeng, A., & Daland, R. (2018). Substantive bias and the acquisition of final (de)voicing patterns. *Paper presented at the 92nd Annual Meeting of the Linguistic Society of America*. Salt Lake City, January 2018.

Greenwood, A. (2016). *An experimental investigation of phonetic naturalness* (Doctoral dissertation). University of California, Santa Cruz.

Hansson, G. Ó. (2008). Diachronic explanations of sound patterns. *Language and Linguistics Compass, 2*(5), 859–893. DOI: https://doi.org/10.1111/j.1749-818X.2008.00077.x

Hansson, G. Ó. (2010). *Consonant harmony: Long-distance interaction in phonology*. Berkeley, CA: University of California Press. Retrieved from https://escholarship.org/uc/item/2qs7r1mw

Hayes, B. (2004). Phonological acquisition in Optimality Theory. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Fixing priorities: constraints in phonological acquisition* (pp. 158–203). Cambridge: Cambridge University Press.

Hayes, B., & White, J. C. (2013). Phonological naturalness and phonotactic learning. *Linguistic Inquiry, 44*(1), 45–75. DOI: https://doi.org/10.1162/LING_a_00119

Hayward, R. J. (1982). Notes on the Koyra language. *Afrika und Übersee, 65*, 211–268.

Hayward, R. J. (1988). Remarks on the Omotic sibilants. In M. Bechhaus-Gerst, & F. Serzisko (Eds.), *Cushitic-Omotic: Papers from the International Symposium on Cushitic and Omotic languages, Cologne, January 6–9, 1986* (pp. 263–299). Hamburg: Helmut Buske Verlag.

Hayward, R. J. (1990). Notes on the Aari language. In R. J. Hayward (Ed.), *Omotic language studies* (pp. 425–493). London: School of Oriental and African Studies, University of London.

Heinz, J. (2010). Learning long-distance phonotactics. *Linguistic Inquiry, 41*(4), 623–661. DOI: https://doi.org/10.1162/LING_a_00015

Heinz, J., & Idsardi, W. J. (2011). Sentence and word complexity. *Science, 333*, 295–297. DOI: https://doi.org/10.1126/science.1210358

Heinz, J., Rawal, C., & Tanner, H. G. (2011). Tier-based strictly local constraints for phonology. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics* (pp. 58–64). Portland, OR: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=2002736.2002750

Hyman, L. M. (1976). Phonologization. In A. Juilland (Ed.), *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday. Vol. 2: Phonology* (pp. 407–418). Saratoga, NY: Anma Libri.

Jäger, G., & Rogers, J. (2012). Formal language theory: Refining the Chomsky hierarchy. *Philosophical Transactions of the Royal Society B, 367*, 1956–1970. DOI: https://doi.org/10.1098/rstb

Jardine, A., & Heinz, J. (2016). Learning Tier-based Strictly 2-Local languages. *Transactions of the Association for Computational Linguistics, 4*, 87–98. DOI: https://doi.org/10.1162/tacl_a_00085

Jardine, A., & McMullin, K. (2017). Efficient learning of Tier-based Strictly *k*-Local languages. In F. Drewes, C. Martín-Vide, & B. Truthe (Eds.), *Proceedings of the 11th*

Language and Automata Theory and Applications conference (LATA 2017) (pp. 64–76). Berlin: Springer. DOI: https://doi.org/10.1007/978-3-319-53733-7_4

Jesney, K., & Tessier, A.-M. (2011). Biases in Harmonic Grammar: The road to restrictive learning. *Natural Language and Linguistic Theory*, *29*, 251–290. DOI: https://doi.org/10.1007/s11049-010-9104-2

Jurgec, P. (2011). *Feature spreading 2.0: A unified theory of assimilation* (Doctoral dissertation). University of Tromsø.

Kapatsinski, V. (2010). Velar palatalization in Russian and artificial grammar: Constraints on models of morphophonology. *Laboratory Phonology*, *1*(2), 361–393. DOI: https://doi.org/10.1515/labphon.2010.019

Kazazis, K. (1969). Possible evidence for (near-)underlying forms in the speech of a child. In R. I. Binnick, A. Davidson, G. M. Green, & J. L. Morgan (Eds.), *Papers form the 5th regional meeting of the Chicago Linguistic Society* (pp. 382–388). Chicago, IL: Chicago Linguistic Society.

Kiparsky, P. (1968). Linguistic universals and linguistic change. In E. Bach, & R. Harms (Eds.), *Universals in linguistic theory* (pp. 170–202). New York: Holt, Rinehart and Winston.

Kiparsky, P. (2008). Universals constrain change; change results in typological generalizations. In J. Good (Ed.), *Linguistic universals and linguistic change* (pp. 23–53). Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199298495.003.0002

Kiparsky, P. (2015). Phonologization. In P. Honeybone, & J. C. Salmons (Eds.), *The Oxford handbook of historical phonology* (pp. 563–579). Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780199232819.013.017

Kuo, L.-J. (2009). The role of natural class features in the acquisition of phonotactic regularities. *Journal of Psycholinguistic Research*, *38*(2), 129–150. DOI: https://doi.org/10.1007/s10936-008-9090-2

Lai, R. (2012). *Domain specificity in learning phonology* (Doctoral dissertation). University of Delaware.

Lai, R. (2015). Learnable vs. unlearnable harmony patterns. *Linguistic Inquiry*, *46*(3), 425–451. DOI: https://doi.org/10.1162/LING_a_00188

Linzen, T., & Gallagher, G. (2017). Rapid generalization in phonotactic learning. *Laboratory phonology*, *8*(1), 1–32. DOI: https://doi.org/10.5334/labphon.44

Luo, H. (2017). Long-distance consonant agreement and subsequentiality. *Glossa*, *52*(1), 1–25. DOI: https://doi.org/10.5334/gjgl.42

McCarthy, J. J. (1998). Morpheme structure constraints and paradigm occultation. In M. C. Gruber, D. Higgins, K. Olson, & T. Wysocki (Eds.), *Papers from the 32nd meeting of the Chicago Linguistic Society*, *2*, 123–150. Chicago, IL: Chicago Linguistic Society.

McMullin, K. (2016). *Tier-based locality in long-distance phonotactics: Learnability and typology* (Doctoral dissertation). University of British Columbia.

McMullin, K., Aksënova, A., & De Santo, A. (2019). Learning phonotactic restrictions on multiple tiers. In *Proceedings of the Society for Computation in Linguistics*, *2*(55). DOI: https://doi.org/10.7275/s8ym-bx57

McMullin, K., & Hansson, G. Ó. (2014). Locality in long-distance phonotactics: Evidence for modular learning. In J. Iyer, & L. Kusmer (Eds.), *Proceedings of the 44th meeting of the North Eastern Linguistic Society*, *2*, 1–14. Amherst: GLSA Publications, University of Massachusetts.

McMullin, K., & Hansson, G. Ó. (2015). Long-distance phonotactics as Tier-based Strictly 2-Local languages. In A. Albright, & M. Fullwood (Eds.), *Proceedings of the 2014 Annual Meeting on Phonology*. Washington, DC: Linguistic Society of America. DOI: https://doi.org/10.3765/amp.v2i0.3750

McMullin, K., & Hansson, G. Ó. (2016). Computational and learnability properties of conflicting long-distance dependencies. *Paper presented at the 24th Annual Manchester Phonology Meeting.* University of Manchester.

McNaughton, R., & Papert, S. A. (1971). *Counter-free automata.* Cambridge, MA: MIT Press.

Mitchell, T. M. (1990). The need for biases in learning generalizations. In J. W. Shavlik, & T. G. Dietterich (Eds.), *Readings in machine learning* (pp. 184–191). San Mateo, CA: Morgan Kaufmann.

Mohri, M. (1997). Finite-state transducers in language and speech processing. *Computational Linguistics, 23,* 269–311. Retrieved from http://dl.acm.org/citation.cfm?id=972695.972698

Moreton, E. (2008). Analytic bias and phonological typology. *Phonology, 25,* 83–127. DOI: https://doi.org/10.1017/S0952675708001413

Moreton, E. (2012). Inter- and intra-dimensional dependencies in implicit phonotactic learning. *Journal of Memory and Language, 67,* 165–183. DOI: https://doi.org/10.1016/j.jml.2011.12.003

Moreton, E., & Pater, J. (2012a). Structure and substance in artificial phonology learning. Part 1: Structure. *Language and Linguistics Compass, 6*(11), 686–701. DOI: https://doi.org/10.1002/lnc3.363

Moreton, E., & Pater, J. (2012b). Structure and substance in artificial phonology learning. Part 2: Substance. *Language and Linguistics Compass, 6*(11), 702–718. DOI: https://doi.org/10.1002/lnc3.366

Myers, S., & Padgett, J. (2014). Domain generalisation in artificial language learning. *Phonology, 31*(3), 399–433. DOI: https://doi.org/10.1017/S0952675714000207

Odden, D. (1994). Adjacency parameters in phonology. *Language, 70,* 289–330. DOI: https://doi.org/10.2307/415830

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. Mac-Neilage (Ed.), *The production of speech* (pp. 189–216). New York: Springer. DOI: https://doi.org/10.1007/978-1-4613-8202-7_9

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: problems and perspectives* (pp. 237–278). London: Longman.

Ohala, J. J. (1994). Towards a universal, phonetically-based, theory of vowel harmony. In *Proceedings of the 3rd International Conference on Spoken Language Processing (ICSLP 3), 2,* 491–494. Yokohama: Acoustical Society of Japan.

Pater, J. (2009). Weighted constraints in generative linguistics. *Cognitive Science, 33*(6), 999–1035. DOI: https://doi.org/10.1111/j.1551-6709.2009.01047.x

Payne, A. (2014). Dissimilation as a subsequential process. In J. Iyer, & L. Kusmer (Eds.), *Proceedings of the 44th meeting of the North Eastern Linguistic Society, 1,* 79–90. Amherst: GLSA Publications, University of Massachusetts.

Peperkamp, S., & Dupoux, E. (2007). Learning the mapping from surface to underlying representations in an artificial language. In J. Cole, & J. I. Hualde (Eds.), *Laboratory Phonology 9,* 315–338. Berlin: Mouton de Gruyter.

Piggott, G. L. (1996). Implications of consonant nasalization for a theory of harmony. *Canadian Journal of Linguistics, 41,* 141–174. DOI: https://doi.org/10.1017/S0008413100016303

Pinker, S. (1979). Formal models of language learning. *Cognition, 7,* 217–283. DOI: https://doi.org/10.1016/0010-0277(79)90001-5

Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar.* Malden, MA: Blackwell. (Originally distributed 1993 as Technical Report RuCCS-TR-2/CU-CS-696-93, Rutgers Center for Cognitive Science, Rutgers University,

and Department of Cognitive Science, University of Colorado at Boulder). DOI: https://doi.org/10.1002/9780470759400

Prince, A., & Tesar, B. (2004). Learning phonotactic distributions. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Constraints in phonological acquisition* (pp. 245–291). Cambridge: Cambridge University Press.

Pulleyblank, D. (2003). Harmony drivers: No disagreement allowed. In J. Larson, & M. Paster (Eds.), *Proceedings of the 28th annual meeting of the Berkeley Linguistics Society* (pp. 249–267). Berkeley, CA: Berkeley Linguistics Society. DOI: https://doi.org/10.3765/bls.v28i1.3841

Pycha, A., Nowak, P., Shin, E., & Shosted, R. (2003). Phonological rule-learning and its implications for a theory of vowel harmony. In G. Garding, & M. Tsujimura (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics* (pp. 423–435). Somerville, MA: Cascadilla Press.

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna. Retrieved from https://www.R-project.org/

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855–863. DOI: https://doi.org/10.1016/S0022-5371(67)80149-X

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*, 219–235. DOI: https://doi.org/10.1037/0096-3445.118.3.219

Rogers, J., & Pullum, G. K. (2011). Aural pattern recognition experiments and the subregular hierarchy. *Journal of Logic, Language and Information, 20*(3), 329–342. DOI: https://doi.org/10.1007/s10849-011-9140-2

Rose, S., & Walker, R. (2004). A typology of consonant agreement as correspondence. *Language, 80*, 475–531. DOI: https://doi.org/10.1353/lan.2004.0144

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry, 21*, 720–731. Retrieved from https://www.jstor.org/stable/4178959

Stanton, J. (2018). Sundanese [ar~al] allomorphy as aggressive reduplication. *Paper presented at the GLOW Workshop on Long-Distance Segmental Phenomena*, Budapest, April 10, 2018.

Staubs, R., Becker, M., Potts, C., Pratt, P., McCarthy, J. J., & Pater, J. (2010). *OT-Help 2.0. Software package.* Amherst, MA: University of Massachusetts Amherst. Retrieved from http://people.umass.edu/othelp/

Suzuki, K. (1998). *A typological investigation of dissimilation* (Doctoral dissertation). University of Arizona.

Walker, R. (2000). Yaka nasal harmony: Spreading or segmental correspondence? In L. Conathan, J. Good, D. Kavitskaya, A. Wulf, & A. C. L. Yu (Eds.), *Proceedings of the 26th annual meeting of the Berkeley Linguistics Society* (pp. 321–332). Berkeley, CA: Berkeley Linguistics Society. DOI: https://doi.org/10.3765/bls.v26i1.1164

Warker, J. A., Xu, Y., Dell, G. S., & Fisher, C. (2009). Speech errors reflect the phonotactic constraints in recently spoken syllables, but not in recently heard syllables. *Cognition, 112*, 81–96. DOI: https://doi.org/10.1016/j.cognition.2009.03.009

Wayment, A. (2009). *Assimilation as attraction: Computing distance, similarity, and locality in phonology* (Doctoral dissertation). Baltimore, MD: Johns Hopkins University.

White, J. (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition, 130*(1), 96–115. DOI: https://doi.org/10.1016/j.cognition.2013.09.008

White, J. (2017). Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a P-map bias. *Language, 93*(1), 1–36. DOI: https://doi.org/10.1353/lan.2017.0001

Wilson, C. (2003). Experimental investigation of phonological naturalness. In G. Garding, & M. Tsujimura (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics* (pp. 533–546). Somerville, MA: Cascadilla Press.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*, 945–982. DOI: https://doi.org/10.1207/s15516709cog0000_89

Yu, A. C. L. (Ed.) (2013). *Origins of sound change: Approaches to phonologization*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199573745.001.0001

Zuraw, K. (2002). Aggressive reduplication. *Phonology*, *19*(3), 395–439. DOI: https://doi.org/10.1017/S095267570300441X