DISTINGUISHING COGNITIVE FROM HISTORICAL INFLUENCES IN PHONOLOGY

GAŠPER BEGUŠ

University of California, Berkeley

Distinguishing cognitive influences from historical influences on human behavior has long been a disputed topic in behavioral sciences, including linguistics. The discussion is often complicated due to empirical evidence being consistent with both the cognitive and the historical approach. This article argues that phonology offers a unique test case for distinguishing historical and cognitive influences on grammar, and it proposes an experimental technique for testing the cognitive factor which controls for the historical factor. The article outlines a model called CATALYSIS for explaining how learnability influences phonological typology and presents experiments that simulate this process. Central to this discussion are unnatural phonological processes, that is, those that operate against universal phonetic tendencies and require complex historical trajectories in order to arise. By using statistical methods for estimating historical influences, mismatches in predictions between the cognitive and historical approaches to typology can be identified. By conducting artificial grammar learning experiments on processes for which the historical approach makes predictions that differ from those of the cognitive approach, the experimental technique proposed in this article controls for historical influences while testing cognitive factors. Results of online and fieldwork experiments on two languages, English and Slovenian, show that subjects prefer postnasal devoicing over postnasal fricative occlusion and devoicing in at least a subset of places of articulation, which aligns with the observed typology. The advantage of the proposed approach over existing experimental work is that it experimentally confirms a link between synchronic preferences and typology that is most likely not influenced by historical biases. Results suggest that complexity avoidance is the primary influence cognitive bias has on phonological systems in human languages. Applying this technique to further alternations should yield new information about those cognitive properties of phonological grammar that are not conflated with historical influences.*

Keywords: cognitive influences, historical bias, phonology, artificial grammar learning experiments, experimental fieldwork, sound change

1. INTRODUCTION. Distinguishing historical (also called cultural or emergent) from cognitive (also called innate) influences on human behavior has long been a topic of discussion in any discipline dealing with human cognition, ranging from psychology to musicology (Altarriba 1993, Gauvain 1995, Nisbett & Norenzayan 2002, Cross 2012). The equivalent dichotomy in linguistics (Kirby et al. 2007, Griffiths et al. 2008, Reali & Griffiths 2009, Kirby et al. 2014, Haynie & Bowern 2016, Ferdinand et al. 2019) and, specifically, phonology (de Lacy 2006b, de Lacy & Kingston 2013) is the distinction between approaches that explain recurrent patterns in the sound systems and phonological alternations of the world's languages primarily from synchronic grammatical constraints or learnability ('cognitive bias' henceforth; Kiparsky 1995, 2006, 2008, Wilson 2006, Moreton 2008, Finley & Badecker 2009, Hayes et al. 2009, Becker et al. 2011, Baer-Henney & van de Vijver 2012, Finley 2012, Moreton & Pater 2012a,b) and ap-

* The experiments reported here were funded by the Mind Brain Behavior Interfaculty Initiative and the Department of Linguistics at Harvard University, as well as by startup funds for new faculty at the University of Washington. I would like to thank Kevin Ryan, Jay Jasanoff, Adam Albright, and Donca Steriade, as well as the editors and referees at *Language*, for useful comments on previous versions of this work. Special thanks go to the nine undergraduate research assistants at the University of Ljubljana who assisted with field experiments: Manca Uršič, Dafne Marko, Julija Kos, Karen Kavčič, Nina Kocjančič, Haris Agović, Denis Hacin, Ariela Herček, and Luka Horjak. I am grateful to Tatjana Marvin and Ana Mehle, who helped coordinate the undergraduate research assistants, and to Samo Beguš, who kindly allowed me to use the anechoic room at the Faculty of Electrotechnics, University of Ljubljana. All mistakes are my own.

proaches that explain these recurrent patterns as emergent from the historical transmission of language in speech communities in time and space ('historical bias' henceforth; Hyman 1975, Greenberg 1978, Ohala 1981, 1983, 1993, Blevins 2004, Hansson 2008, Morley 2012, Garrett & Johnson 2013).[1]

A phonological process that exemplifies the discussion around the historical versus cognitive pressures in phonology is postnasal voice alternation. Postnasal voicing is a phonological alternation in which voiceless stops /p, t, k/ are realized as the corresponding voiced stops [b, d, g] when they appear after nasal sounds [m, n, ŋ] (represented as T → D / N _).[2] Greek /to**n** **t**opo/ 'the place' surfaces as [to**nd**opo], because /t/ appears after a nasal and is realized as a voiced [d] (Pater 2004). Its opposite process is POSTNASAL DEVOICING (PND; represented as D → T / N _), where voiced stops /b, d, g/ are realized as voiceless stops [p, t, k] after nasal sounds [m, n, ŋ]; for example, in Shekgalagari /χʊ**mb**ɔ́ná/ 'to see me' is realized as [χʊ**mp**ɔ́ná] (Solé et al. 2010). Distribution of phonological alternations across the world's languages is not uniform—some alternations are substantially more frequent than others. Postnasal voicing, for example, is a comparatively widespread alternation, occurring in approximately twenty-eight of 629 languages surveyed (Mielke 2018). Postnasal devoicing, by contrast, is comparatively rare as a productive alternation, occurring in only two or three languages among the approximately 600 surveyed (Hyman 2001, Beguš 2019). Distributional asymmetries like these highlight the debate about whether these asymmetries are driven by cognitive or historical factors.

Discussing cognitive and historical influences in linguistics is complicated both by terminology and by matters of substance (§1.1). Here, I attempt to clarify several concepts that enter the discussion. I describe with the term COGNITIVE BIAS any influence of both domain-specific grammatical and domain-general cognitive mechanisms that result in typological asymmetries and operate as synchronic tendencies in individual speakers. While the line between domain-specific grammatical and domain-general cognitive influences is sometimes blurred, it is possible to distinguish between the two in some cases. Cognitive-bias influences have long been divided into SUBSTANTIVE BIAS and COMPLEXITY BIAS. Substantive bias states that phonetically unmotivated processes are dispreferred by the grammar compared to phonetically motivated processes, and are, as such, predicted to be less frequent. Complexity bias, by contrast, states that complex alternations (e.g. those involving more features or those involving perceptually more distant sounds) are dispreferred or more difficult to learn. Complexity bias has been confirmed in numerous studies (see Moreton & Pater 2012a); complex alternations are therefore predicted to be typologically less frequent (for ambiguous experimental outcomes regarding substantive bias, see §1.1).

Domain-specific grammatical and domain-general cognitive influences occasionally align with substantive and complexity biases. For example, there is no clear cognitive domain-general reason why the unnatural D → T / N _ (e.g. in /χʊ**mb**ɔ́ná/ → [χʊ**mp**ɔ́ná]) would be dispreferred compared to its exact opposite process T → D / N _ (e.g. /to**n** **t**opo/ → [to**nd**opo]). Under the influential proposal of PHONETICALLY BASED PHONOLOGY (Hayes 1999, Hayes & Steriade 2004), grammatical constraints are phonetically grounded, so ultimately it is the phonological GRAMMAR that makes the unnat-

[1] Various alternative names for the two approaches exist in the literature: the cognitive bias is also known as 'analytic' or 'learnability' bias, the historical bias as the 'channel' bias (Moreton 2008).

[2] The following capital letters represent different groups of segments: N: nasals, such as [m, n, ŋ]; T: voiceless stops, such as [p, t, k]; D: voiced stops, such as [b, d, g]; S: voiceless fricatives, such as [f, θ, x]; Z: voiced fricatives, such as [v, ð, ɣ]; C: consonant; V: vowel. All symbols representing sounds of language follow International Phonetic Alphabet conventions.

ural alternations dispreferred under this proposal. For example, postnasal devoicing is dispreferred by the substantive bias because it operates against phonetic naturalness, but it is not more complex than the natural postnasal voicing. In the absence of a domain-general cognitive explanation, this particular dispreference can be understood as a domain-specific bias. By contrast, dispreference for complex alternations that target more than one feature and also involve perceptually distant allophones, such as postnasal fricative occlusion and devoicing (PFOD; Z → T / N _ , e.g. /**mß**ona/ → [**mp**ona] in Pedi; Dickens 1984) can be explained with domain-general cognitive mechanisms. Moreton et al. (2017) adopt concept learning and argue that complexity dispreference in phonology might have the same underlying mechanisms as other cognitive processes, where concepts that require more features to be described are more difficult to learn.

These grammatical dispreferences for phonetically unnatural (substantive bias) or complex processes (complexity bias) can result from several mechanisms. For example, learning asymmetries (Wilson 2006) can underlie both complexity and substantive bias. Featurally complex concepts are difficult to learn not only in phonology, but also in other domains (Moreton et al. 2017), which would explain the relative rarity of complex alternations. However, it has also been assumed that phonetically unnatural processes are difficult to learn; thus, learning difficulties are also the basis of the substantive bias. Another mechanism that can underlie the substantive bias is the inability of the grammar to accommodate an unnatural process (e.g. as a tendency to reanalyze an unnatural process as a process that conforms to naturalness; Kiparsky 2006, 2008).

Finally, perceptual forces can influence typology. Whether perceptual influences should be analyzed as part of complexity or substantive bias is an open question. While this distinction is primarily terminological in nature and does not crucially affect the results reported in this article, I analyze perceptual forces as part of the complexity bias. First, perceptual distance can often be directly analyzed in terms of formal featural complexity. Even if two alternations are equally complex featurally, the tendency toward minimizing some distance can be understood as part of complexity (somewhat diverging from the literature in Wilson 2006 and White 2014): similarity is less complex than dissimilarity. Finally, while the substantive bias has traditionally been assumed to be limited to domain-specific processes, the preference for perceptual similarity is not domain-specific, but likely a domain-general cognitive mechanism (in the visual domain, see e.g. Schloss & Palmer 2011). A tendency to keep phonology perceptually minimal (the P-map; Steriade 2001) can thus be analyzed as part of the complexity bias, while substantive bias is reserved for a domain-specific dispreference for phonetically unmotivated or unnatural processes.

With historical bias, by contrast, I describe those properties that emerge when articulatory and perceptual tendencies in speech production and perception operate in language use as language is transmitted in space and time across generations of speakers and which accumulate in phonological or phonetic processes. Production of speech is a highly variable process, and variation in speech production is biased. For example, stops [p, t, k] tend to be produced with a higher degree of voicing (more like [b, d, g]) if they appear after nasal stops [n, m, ŋ] (Hayes & Stivers 2000, Davidson 2016, 2018). Such minor biased phonetic variation, motivated by articulatory or perceptual factors, gives rise to phonological alternations via sound change and the process of phonologization (Hyman 2013).

Articulatory and perceptual forces differ in terms of their relatedness to cognitive influences. There is little cognitive influence on articulatory forces: automatic articulatory tendencies are often motivated by motor-planning mechanisms dissociated from higher-level cognitive processes. For example, the motivation for postnasal voicing (/to**n** **t**opo/ → [to**nd**opo]) is purely mechanical: when the velum rises to close the nasal cavity (from [n] to [t]), the volume of the oral cavity increases, and airflow can continue for some period of time as the velum is rising. Because increased volume and airflow promote voicing, stops in postnasal position feature more voicing into the closure than those in other positions (Hayes & Stivers 2000, Coetzee & Pretorius 2010, Davidson 2016, 2018). These articulatory mechanisms are purely mechanical influences and have little connection to cognition. However, the historical bias includes perceptual influences as well. Perception is not dissociated from cognition, but perceptual influences from the historical-bias perspective are distinct from synchronic perceptual influences. First, the perceptual mechanisms in Ohala's (1981) terms result in sound change by operating gradually in a speech community in space and time (i.e. historical-bias influences). Second, perceptual mechanisms often (but not always) require some minor phonetic variation that originates in noncognitive articulatory forces. For example, longer vowel duration before voiced (vs. voiceless) stops is likely caused by perceptual enhancement (Kluender et al. 1988), but the initial distribution on which the perceptual forces operate likely originates in articulatory factors (Beguš 2017).

In sum, perceptual influences can be part of both cognitive and historical biases. It is possible that perceptual forces result in typological patterns because of cognitive synchronic preferences for phonological alternations to be minimally distant in terms of perception (the P-map; Steriade 2001) or because of hypo- and hypercorrection (Ohala 1981). The advantage of the proposal advanced here is that we can disambiguate between the two by comparing experimental results against independent historical samples and against those samples in which potential perceptual forces are related to a synchronic phonological alternation. For example, Ohala's (1981) perceptual forces should operate at equal rates in systems in which the target segments are not part of a synchronic alternation, as well as in those in which they are. Because the results here suggest the opposite to be true, we can argue that, even if the observed results are part of perception, they are driven by cognitive factors, not perceptual mechanisms operating in space and time (see §2.3).

The impact of these factors on phonological typology is central to phonology and linguistics in general and has far-reaching consequences. It is likely that both influences affect phonological typology (Hyman 2001, Myers 2002, Moreton 2008, Moreton & Pater 2012a,b, de Lacy & Kingston 2013), but to distinguish the two has been a challenging task, primarily because empirical evidence tends to support both approaches equally well (for nonexperimental attempts, see de Lacy 2006b, de Lacy & Kingston 2013). Distinguishing cognitive factors from those aspects of phonology that are emergent from the historical transmission of language in space and time is a desirable goal: it would yield a better understanding of which properties of phonology—and consequently of human language capacity—are influenced by cognitive processes and should therefore be captured by models of grammar.

1.1. The duplication problem. Empirical research in the discussion of cognitive vs. historical forces in phonology is complicated by several confounding factors: (i) ambiguous empirical evidence, (ii) difficulty in disassociating cognitive influence from phonetic variation, and (iii) lack of elaborate models that link learning and typology.

First, empirical evidence often supports both approaches equally well. Numerous experimental studies have established a link between learnability and typological distribution: on the one hand, phonological alternations that are more difficult to learn in laboratory settings are typologically less frequent (for an overview, see Moreton & Pater 2012a,b). On the other hand, phonetic variation, motivated by articulatory or perceptual mechanisms, has been shown to result in active phonological alternations via the process of phonologization. The stronger the phonetic tendency, the more frequent the resulting phonological alternation (Blevins 2004). Crucially, even if learnability differences that match the observed typology are experimentally confirmed, the typological distribution can nevertheless be explained within the historical-bias approach. Experimental studies testing learnability only rarely target predictions that cannot be attributed to historical bias (for one method, see Moreton 2008, and see Yu 2011 for discussion of some of its shortcomings).

For example, high frequency of postnasal voicing (/to**n t**opo/ → [to**nd**opo]) and low frequency of postnasal devoicing (/χʊ**mb**ɔ́ná/ → [χʊ**mp**ɔ́ná]) can be explained by grammatical or learnability dispreferences for a phonetically unmotivated process. Some proposals even argue that the grammar is incapable of accommodating phonetically unnatural processes (Kiparsky 2006, 2008). The same typological asymmetry can be explained under the historical-bias approach: phonetic variation caused by mechanical factors is present in the former but absent in the latter, which is why sound change can produce postnasal voicing but not devoicing (unless a particular combination of at least three sound changes conspires to produce the unnatural result; see §2.2).

This duplication of evidence goes even further: complexity bias also has a duplicate historical explanation. Complex alternations such as PFOD (/**mβ**ona/ → [**mp**ona]) can be rare because they are more difficult to learn, computationally more complex, or require a large perceptual distance between the target and result (all cognitive biases), or they can be rare because complex alternations require two sound changes, and the occurrence of two historical events (i.e. sound changes) generally has a lower probability than the occurrence of a single event (historical bias) (Bell 1970, 1971, Greenberg 1978, Cathcart 2015, Morley 2015, Beguš 2019).

While experimental results consistent with substantive bias have been reported in some studies (Wilson 2006, Carpenter 2010), many other experimental studies have failed to find positive evidence for substantive bias (Pycha et al. 2003, Kuo 2009, Skoruppa & Peperkamp 2011 via Seidl et al. 2007 and Moreton & Pater 2012a,b, Do et al. 2016, Glewwe 2017, Glewwe et al. 2017, Do & Havenhill 2021). As already mentioned, historical bias makes exactly the same predictions as the complexity- and substantive-bias approaches. Featurally complex alternations are predicted to be less frequent than simple alternations because they result from multiple sound changes (for a discussion, see §2.1). Phonetically motivated alternations are likewise predicted to be more frequent by the historical-bias approach: they arise from a single sound change, while unnatural alternations require a combination of sound changes. This means that experimental results confirming learning biases for a given process almost always have an alternative historical explanation. This is called the DUPLICATION PROBLEM henceforth.

The second confounding factor is that it is possible that the frequency of sound change itself is crucially affected not only by the robustness of phonetic variation (as assumed by the historical-bias approach) but by learnability as well. In other words, what the historical-bias approach assumes to be exclusively an emergent factor—that is, frequency of a sound change based on historical factors such as articulatory or perceptual motivation of phonetic variation—can be influenced by learnability or other cognitive factors (Kiparsky 1995, 2006, 2008, Moreton 2008). For example, Moreton (2008) argues that cognitive factors determine when some phonetic variation will result in a sound change/synchronic alternation: phonetic variation of equal size can result in sound change or not. Cognitive biases are responsible for this asymmetry (for problematic aspects of the notion of equal size in phonetic variation/precursors, see Yu 2011).

Finally, elaborate models of how exactly learning influences the observed typology are lacking. The cognitive-bias approach often uses first language (L1) acquisition to explain the link between learnability and typology, whereby learners fail to learn a process or restructure their phonological grammar based on learning biases. However, many experimental studies have shown that, given enough exposure, any alternation can be learned (Hayes et al. 2009, Coetzee & Pretorius 2010, Hayes & White 2013, White 2014, Avcu 2018). Additionally, human L1 learners get even more exposure to primary linguistic data than subjects in laboratory experiments do, and they are able to reproduce language input with a high degree of faithfulness past some developmental stage (Kong et al. 2012). Numerous studies have also confirmed the anti-alternation bias: learners prefer no alternation to any alternation (Wilson 2006, Tessier 2012, and the literature therein). The fact that sound change gives rise to active phonological alternations means that learnability alone does not crucially affect the operation of contextually limited sound changes that result in alternations. Showing how learnability differences affect the typology is thus not trivial.

The position that anything can be learned given enough exposure is supported by phonological data as well: unnatural alternations operating in the exact opposite direction from universal phonetic tendencies that result from a combination of sound changes are attested as productive alternations in languages such as Tswana and Shekgalagari (postnasal devoicing; Coetzee & Pretorius 2010). This means that learners were able to learn a phonological grammar with an alternation as phonetically unnatural as postnasal devoicing. Second, while L1 acquisition is a potential source of sound change, it is known that sound change operates within early adolescent and adult populations as well (Labov 1994, 2001, Sankoff & Blondeau 2007, and the literature therein). An attempt to derive typology through L1 learning thus needs to account for sound change in the adult population. Third, that phonological processes in L1 acquisition often differ from sound change typology is an (often overlooked and underresearched) observation that poses a challenge to the L1 approach to sound change (Bybee 2001). Finally, some computational models suggest that learnability differences might not be sufficient to derive surface typology (Rafferty et al. 2013).

In sum, it is nontrivial to show how cognitive biases result in phonological typology. To address this difficulty, I propose a model called CATALYSIS, which outlines a possible mechanism by which synchronic cognitive bias can accelerate the operation of a sound change and consequently directly influence the typology. Simulations of catalysis are tested experimentally. Because we can argue that, based on the estimation of historical probabilities (§2.2) behind phonological processes, historical bias makes predictions opposite of those made by cognitive bias, we can dissociate the two influences and test them against the observed typology.

1.2. TESTING THE HYPOTHESES ON MISMATCHED PREDICTIONS. This article argues that phonology offers a unique test case for the debate on cognitive vs. historical influences, one that avoids the duplication problem outlined above if we adopt some assumptions proposed in Beguš 2019. Key to this discussion are unnatural phonological processes, defined in Beguš 2019 as those that operate in exactly the opposite direction from universal phonetic tendencies (such as final voicing or postnasal devoicing). Beguš 2019 argues that sound change cannot operate in an unnatural direction and that unnatural alternations require at least three sound changes to arise (the so-called MINIMAL SOUND change requirement; §2.1). This crucial condition makes unnatural alternations the best testing ground for distinguishing cognitive from historical influences on typology. Even if learnability influences the operation of individual sound changes, the requirement that three historical events (i.e. sound changes) need to operate in a speech community within a given timeframe in order to produce an unnatural alternation can be exclusively ascribed to the influence of the historical factor.

Additionally, the historical- and cognitive-bias approaches make opposing predictions about the relative frequencies of unnatural and complex alternations. These crucial mismatches allow us to test the influences of one approach while controlling for the other. A statistical technique proposed in Beguš 2020 for estimating the probabilities of sound changes based on diachronic factors facilitates the identification of mismatches in predictions between the two approaches to typology.

This article presents experiments that simulate a development from a complex to an unnatural process, and thus experimentally tests the mismatched predictions between the historical and cognitive approach. The historical trajectory required for an unnatural alternation, postnasal devoicing, is first identified, and the historical probabilities of each stage in its development are estimated (based on Beguš 2019, 2020). The experiments test learning of the last two stages in the historical development of postnasal devoicing. The cognitive- and historical-bias approaches make opposing predictions for these stages. The historical-bias approach predicts the unnatural stage (postnasal devoicing, e.g. /b/ → [p] / m _) to be less frequent than the complex stage (postnasal fricative occlusion and devoicing, e.g. /v/ → [p] / m _), while the cognitive-bias approach predicts the opposite. If experimental results support the cognitive-bias approach, which matches the observed typology, the link between cognitive bias and typology is supported without the duplication problem—historical bias is likely not responsible for a typological distribution that operates against its predictions.

The experiments were conducted within the artificial grammar learning paradigm (Albright & Do 2019; see overview in Moreton & Pater 2012a,b), but several additional factors were introduced. To diversify the design, online experiments targeting nonspecific groups of speakers were combined with experimental fieldwork with a high number of subjects and a relatively high train-to-test ratio of stimuli. Supervision of one of the experiments by a research assistant at least partly addresses the concern of subjects' attention in online experiments. Additionally, the experiments were conducted on L1 speakers of English and Slovenian, two languages with different realizations of the feature [±voice] and different frequencies of the segments tested in the experiment, which at least partially controls for interference from first language, a long-standing objection to the artificial grammar learning experimental paradigm.

The results suggest that when presented with ambiguous stimuli, subjects prefer the response consistent with the unnatural alternation compared to the complex alternation, for at least a subset of places of articulation. These results are in line with the observed synchronic typology, and they operate against what is predicted by statistical modeling of the historical bias. This means that the experimentally confirmed link between cognitive bias and the observed typology is likely not due to the historical bias.

2. Background: unnatural phonology. Unnatural phonological alternations are defined as those that operate against universal phonetic tendencies. Universal phonetic tendencies are defined as articulatorily or perceptually motivated (Garrett & Johnson 2013) phonetic processes that 'passively operate in speech production cross-linguistically and result in typologically common phonological processes' (Beguš 2019:691).

Based on a typological and historical study, Beguš 2019 argues that unnatural segmental alternations always arise from a specific combination of three sound changes called the BLURRING PROCESS: (i) a sound change that creates a complementary distribution, (ii) a sound change that targets a subset of segments in the complementary distribution, and (iii) a sound change that undoes the original complementary distribution (for a schematic representation, see Table 1). The 'blurring process' proposal argues that postnasal devoicing results from a combination of three sound changes in all surveyed cases, a crucial assumption adopted from Beguš 2019.[3] First, voiced stops [b, d, g] fricativize to [v, ð, ɣ] except postnasally (D > Z / [−nas] _ ,[4] e.g. [**b**am**b**a] > [**v**am**b**a]). Then, voiced stops [b, d, g] devoice to [p, t, k] unconditionally (D > T, e.g. [vam**b**a] > [vam**p**a]), but because at this stage they surface only postnasally, the resulting devoicing appears to be limited to postnasal position.[5] Finally, voiced fricatives [v, ð, ɣ] occlude back to stops [b, d, g] (Z > D, e.g. [**v**ampa] > [**b**ampa]), which results in a synchronic PND (D → T / N _ or /bam**b**a/ → [bam**p**a]; Table 1).

	BLURRING CYCLE	PND	SCHEMATIC EXAMPLE		
1.	B > C / ¬X	D > Z / [−nas] _	[**b**am**b**a]	>	[**v**am**b**a]
2.	B > A	D > T	[vam**b**a]	>	[vam**p**a]
3.	C > B	Z > D	[**v**ampa]	>	[**b**ampa]
RESULT	B → A / X	D → T / [+nas] _	/bam**b**a/	→	[bam**p**a]

TABLE 1. Blurring cycle (schematic; left) yielding PND (right) (from Beguš 2020).

2.1. MINIMAL SOUND CHANGE REQUIREMENT. Unnatural processes are crucial for identifying mismatched predictions between cognitive and historical biases because they have well-structured complex histories. Beguš 2019 provides a formal argument for the requirement of at least three sound changes in order for an unnatural alternation to arise. In abstract terms, a set of segments represented by feature matrix A can alternate with a set of segments B in environment X. If B is in phonetic terms universally preferred in X, then A → B / X is a natural alternation. The opposite process, B → A / X, is unnatural, where A is phonetically dispreferred in X. In feature matrix notation, a change of a feature value of ϕ_1 from α to ¬α in environment X given a constant set of other features β, where ¬α is preferred in X, is natural. A change of ¬α to α in X given β, where α is dispreferred in X, is unnatural (schematized in Figure 1).

Natural alternation

$$\begin{array}{ccccc} & A & \rightarrow & B & / \; X \end{array}$$

$$\begin{bmatrix} \phi_1 \\ \vdots \\ \phi_{1+n} \end{bmatrix} \begin{bmatrix} \alpha \\ \vdots \\ \beta \end{bmatrix} \rightarrow \begin{bmatrix} \neg\alpha \\ \vdots \\ \beta \end{bmatrix} \; / \; X$$

Unnatural alternation

$$\begin{array}{ccccc} & B & \rightarrow & A & / \; X \end{array}$$

$$\begin{bmatrix} \phi_1 \\ \vdots \\ \phi_{1+n} \end{bmatrix} \begin{bmatrix} \neg\alpha \\ \vdots \\ \beta \end{bmatrix} \rightarrow \begin{bmatrix} \alpha \\ \vdots \\ \beta \end{bmatrix} \; / \; X$$

FIGURE 1. Natural and unnatural alternations.

[3] To be sure, it cannot be proven that PND cannot operate as a single sound change, but see Beguš 2018, 2019 for arguments that provide direct evidence for intermediate stages and for an example that strongly suggests that the unnatural intervocalic devoicing cannot result from a single sound change.

[4] The feature 'nasal' is abbreviated as [±nas] henceforth.

[5] While postnasal position facilitates voicing into the closure, stop closure is nevertheless antagonistic to voicing, and speakers need to actively adjust for voicing even postnasally. For an extensive discussion on the naturalness of unconditioned stop devoicing when stops appear in postnasal position, see Beguš 2019. There exists phonetic evidence that supports this assumption: English voiced stops are relatively frequently realized as partially voiceless in postnasal position (Davidson 2016, 2018, Beguš 2019) due to the antivoicing effects of closure.

Beguš 2019 assumes that sound change is always natural and cannot produce an unnatural alternation in a single step. It is also assumed (following the MINIMALITY PRINCIPLE in Picard 1994) that sound change almost always targets a single feature value (for a detailed discussion, see Beguš 2019).[6] $\neg\alpha > \alpha$ can thus not operate under a constant set of feature values β in the matrix. To get $\neg\alpha > \alpha$, β first needs to change to $\neg\beta$. Under this new condition, $\neg\alpha > \alpha$ can become phonetically motivated and can operate as a sound change. To get the full unnatural alternation, $\neg\beta$ then has to change back to β. Minimally three independent historical events, that is, sound changes, are thus required for an unnatural process to arise (Figure 2). For details, see Beguš 2019.

$$\begin{array}{cccccccc} \textbf{Stage} & 1. & > & 2. & > & 3. & > & 4. \\ \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_{1+\mathbf{n}} \end{bmatrix} & \begin{bmatrix} \neg\alpha \\ \vdots \\ \beta \end{bmatrix} & & \begin{bmatrix} \neg\alpha \\ \vdots \\ \neg\beta \end{bmatrix} & & \begin{bmatrix} \alpha \\ \vdots \\ \neg\beta \end{bmatrix} & & \begin{bmatrix} \alpha \\ \vdots \\ \beta \end{bmatrix} \end{array}$$

FIGURE 2. Changes in feature values in a blurring process (from Beguš 2019).

2.2. HISTORICAL PROBABILITIES. The historical-bias influences on phonological typology can be quantitatively estimated by combining the concept of alternations requiring a specific number of historical events (i.e. sound changes) with the estimation of individual probabilities of these historical events (Beguš 2019, 2020). In other words, the HISTORICAL PROBABILITY (P_χ) of an alternation, that is, the probability that an alternation arises based on historical factors, can be estimated from the number of sound changes the alternation requires and each change's respective probability (adopted from Beguš 2020).

The probability of each individual sound change S_i (in 1 below) is estimated based on historical typological surveys of sound changes (Kümmel 2007) from the number of occurrences (successes) of a sound change S_i and the number of languages surveyed (successes and failures). 95% confidence intervals adjusted for bias and skewness (BC_a) are estimated using nonparametric bootstrap (Efron 1979, 1987). The historical probability of alternation A_j, which requires more than a single sound change, is estimated as a joint probability (a simple product of each individual change; for the assumption of independence, see Beguš 2020) of individual sound changes corrected for ordering of sound changes ($n!$). The historical probability is again estimated with nonparametric bootstrap (in 2). For all details and underlying assumptions of the model, as well as for a discussion on the representativeness of samples, see Beguš 2020.

(1) $P_\chi(S_i) = \frac{\text{number of languages with sound change } S_i}{\text{number of languages surveyed}}$

(2) $P_\chi(A_j) = \frac{\prod_{i=1}^{n} P_\chi(S_i)}{n!}$

2.3. MISMATCHES. Each stage in the development of PND (the blurring process; Table 1) has a historical probability that can be estimated using the bootstrapping technique (outlined in §2.2 above and in Beguš 2020). Table 2 lists counts of languages with the three sound changes that operate in the development of PND in the historical sample given in Kümmel 2007. The first sound change, D > Z / [−nas]/V _ (V), results in a synchronic alternation between voiced stops, which surface postnasally, and voiced frica-

[6] Most of the predictions in Beguš 2019 still hold even if the minimality principle is a tendency rather than a hard rule.

tives, which surface elsewhere (D → Z / [−nas] _ , e.g. /b/ → [v] / [−nas] _). One phonological feature, [±continuant], is manipulated in this alternation. When the second sound change, unconditioned devoicing of voiced stops (D > T, e.g. [b] > [p]), occurs, the resulting alternation is PFOD: voiced fricatives in the elsewhere condition alternate with voiceless stops postnasally (Z → T / N _ , e.g. /v/ → [p] / N _). This alternation manipulates two phonological features, [±continuant] and [±voice], but the latter is automatic, because at that point, the system lacks voiced stops altogether.

SOUND CHANGE	EXAMPLE	COUNTS	SURVEYED
	[bamba]		
D > Z / [−nas]/V _ (V)	[vamba]	47	294
D > T	[vampa]	15	263
Z > D	[bampa]	17	216

TABLE 2. Counts of languages that feature the sound changes required for PND to arise ('counts') and counts of languages surveyed ('surveyed') for each corresponding sound change (from Beguš 2020). Counts are based on a historical sample in Kümmel 2007.

Finally, the third sound change, occlusion of fricatives (Z > D, e.g. [v] > [b]), results in the unnatural PND where underlying voiced stops surface as voiceless postnasally and as voiced elsewhere. PND manipulates one feature, [±voice]. Table 3 shows estimated HISTORICAL PROBABILITIES (P_χ) (with 95% BC_a confidence intervals) of each of the three stages in the blurring process, which result from operation of the first, the first and second, and all three sound changes.

SOUND CHANGE	ALTERNATION	P_χ	LO.	UP.	FEATURES	P_χ	P_{cplx}
	No alternation	83.5			0		
D > Z / [−nas] _	D → Z / [−nas] _	16.0	11.9	20.1	1	↓	↓
D > T	Z → T / [+nas] _	0.5	0.3	0.8	2	↓	↓
Z > D	PND	0.01	0.006	0.02	1	[↓]	[↑]

TABLE 3. Estimated historical probabilities (P_χ in %) of each sound change, with lower ('lo.') and upper ('up.') 95% BC_a confidence intervals and the number of phonological features the resulting alternation manipulates. Arrows in the last two columns indicate changes in the predicted probability (compared to previous stage) of the historical bias (P_χ) and the complexity bias within the cognitive bias (P_{cplx}); table from Beguš 2020.

A clear mismatch in predictions between historical bias and cognitive bias emerges (Beguš 2020) if it is assumed that a single sound change cannot operate in the phonetically unmotivated direction (e.g. PND is not a possible sound change; for an overview of the literature that holds this view and for detailed argumentation, see Beguš 2019). Given this assumption, historical bias predicts that unnatural alternations are significantly less frequent than natural alternations: the probability of an alternation decreases with each additional sound change, regardless of the complexity of the resulting alternation. The probability of $n + 1$ events is always lower than the probability of n events, and this influence is exclusively the result of the historical bias, as sound changes are historical events operating in speech communities in time and space. By contrast, complexity bias predicts featurally complex alternations to be typologically less frequent than featurally simple alternations (see §1). This prediction works regardless of what the underlying mechanism behind complexity bias is: structural (featural) complexity akin to concept learning (Moreton et al. 2017) or perceptual distance (in terms of the P-map; Steriade 2001). Complexity avoidance, where featurally or perceptually complex processes are tested against simple processes, is experimentally confirmed in numerous studies (Moreton & Pater 2012a). The cognitive-bias approach thus predicts PND to be more frequent than PFOD, because the first requires manipulation of one less feature than is required by the second (or it is perceptually less distant). Conversely, the historical-bias approach predicts the unnatural PND to be less frequent than PFOD, because the first requires an additional sound change (Beguš 2020).

Because no learning differences are observed between the natural postnasal voicing and the unnatural PND (Do et al. 2016, Do & Havenhill 2021), substantive bias alone is not likely to be responsible for the observed typological asymmetries. A clarification is warranted here (from Beguš 2020). The only types of learning differences between the natural and unnatural alternations (substantive bias) are those that involve articulatory effort: segments that require greater articulatory effort are acquired later (Clark & Bowerman 1986, Kong et al. 2012, Broselow 2018), which means that, for example, word-final voiced stops are acquired later in L1 and L2 acquisition. Do and Havenhill (2021) also find limited evidence that exposure to production increases natural responses in the postnasal position in adults, but it is not unexpected that an experiment with production (more exposure) improves learning in the variable condition.[7] The mechanism underlying this articulatory learning is likely different from phonological learning, as observed, for example, in complexity bias (Moreton & Pater 2012a,b), and is, in fact, more consistent with the historical-bias approach. Unnatural alternations require the production of segments requiring more articulatory effort.

That the learning of articulatorily more complex gestures is more difficult is not surprising. The very same mechanism is responsible for universal phonetic tendencies operating in the adult population: the articulatory effort of different gestures causes varying degrees of phonetic variation, which result in phonological alternations via phonologization. This mechanism thus falls within the historical-bias approach (Beguš 2020). Deriving typology from articulatory learning within the cognitive-bias approach remains problematic, as children are able to replicate their linguistic input with a high degree of faithfulness past some developmental age (Kong et al. 2012). One of the main pieces of evidence against L1 articulatory effort influencing typology comes from the observation that many articulatory adjustments in L1 acquisition do not result in sound changes if the variation is also not present in the adult language (Bybee 2001).

In sum, the last sound change in the development of PND (the blurring process) decreases the historical probability of the resulting alternation, but increases its synchronic preference, because it reduces its structural or perceptual complexity. This means that the historical-bias approach predicts the unnatural alternation PND to be less frequent than PFOD, whereas the cognitive-bias approach makes the opposite prediction.

These predictions can be directly tested against the observed typological distributions (Beguš 2020). Historical bias predicts PFOD to be approximately fifty times more typologically frequent than PND is, as estimated based on counts in Table 3 (P_χ of PFOD is 0.5% [0.3%, 0.8%]; P_χ of PND is 0.01% [0.006%, 0.02%]). A typological survey in Beguš 2019 shows that PFOD is indeed more frequent than PND, but not as much as would be predicted by historical bias. The survey in Kümmel 2007 and Beguš 2019 is complemented by the most comprehensive survey of phonological rules, P-base

[7] This is especially so because English postnasal voiceless stops have a considerable amount of voicing into the closure (Davidson 2016, 2018). In other words, English has a gradual phonetic process of postnasal devoicing that operates during production and can thus affect experimental results. Do and Havenhill (2021) report no differences in learning in the categorical condition. Moreover, the test phase in Do & Havenhill 2021 was exclusively orthographic, which can introduce confounds.

(Mielke 2018), where PFOD is reported as a synchronic alternation in only one system (Sie).[8] In Beguš 2020, approximately three to six languages are reported to feature PFOD. PND, by contrast, is confirmed as a productive alternation in two closely related languages (Tswana and Shekgalagari) and is reported, but not yet confirmed, in one additional language (Buginese). In seven languages PND is reported as a sound correspondence resulting from a combination of sound changes. Exact counts are difficult to determine because the productivity of a synchronic alternation needs to be experimentally confirmed, and experimental work on many of these languages is lacking. Based on available typological data, however, PFOD seems to be substantially less than fifty times more frequent than PND (Beguš 2020).

Another, more testable mismatch in predictions emerges from the proposed technique of estimating historical probabilities. The last sound change in the blurring process that leads to PND—the occlusion of voiced fricatives to stops (Z > D, e.g. [v] > [b])—appears to operate more frequently than would be expected by only the historical-bias approach (Beguš 2020). To test this observation, the historical probability of a sound change operating in an unconditioned sample is compared to the historical probability of the same sound change operating on languages that already underwent the first two sound changes in the blurring process (i.e. languages where the sound change would simplify an alternation). The probability of the occlusion of fricatives (Z > D) operating independently (forty-four times in a sample of 216 languages in Kümmel 2007) is compared to the probability of the same sound change when it operates as the last sound change in the blurring process (occlusion targets at least one place of articulation in six of ten languages with PFOD in the sample in Beguš 2020 and in P-base; Mielke 2018). Counts are given in Table 4; for details, see Beguš 2020. Occlusion of fricatives occurs significantly more frequently as the last sound change in the development of PND (where it reduces the complexity of the resulting alternation) than it does in the independent sample ($p < 0.01$, Fisher's exact test; Beguš 2020). In other words, a sound change that simplifies a featurally complex alternation and thereby learnability of the alternation operates significantly more frequently than predicted by the historical-bias approach. While the sample in the condition group is small, the comparison suggests that cognitive bias influences the frequency of sound change in this type of case (Beguš 2020).

	✓	✘	TOTAL
INDEPENDENT	44	172	216
SIMPLIFIES	6	4	10

Fisher's exact test: $p < 0.01$

TABLE 4. Contingency table and Fisher's exact test of counts of occurrence (successes vs. failures) of a sound change in an independent sample and in cases where the sound change simplifies an alternation, based on Kümmel 2007, Mielke 2018, and Beguš 2020.

3. CATALYSIS. As mentioned in §1, elaborate models of how cognitive preferences and typology are connected are lacking, and several objections have been raised against the existing models. Here I propose a possible mechanism for the direct link between cognitive bias and typology.

Phonetic variation resulting from universal articulatory or perceptual phonetic tendencies is the underlying condition for every nonanalogical sound change (see Moreton

[8] The query for searching P-base consisted of a [−stop][+voiced] condition for the search input and a [+stop][−voiced] condition for the search output.

2008, Yu 2011, Garrett & Johnson 2013, and §1). For example, the occlusion of fricatives (Z > D, e.g. [v] > [b]), which operates in the development of PND, is a well-documented sound change (Kümmel 2007) with a relatively well-understood phonetic motivation: articulatory targets for fricatives require greater precision compared to stops (Ladefoged & Maddieson 1996:137). Reducing the precision of articulatory targets can result in the occlusion to stops. Typologically, occlusion of fricatives (targeting at least a subset of places of articulation) is a relatively frequent sound change: it is attested in approximately forty-four of 216 languages surveyed (see Table 4, Beguš 2020, and Kümmel 2007).

Occlusion of fricatives can thus operate as a passive phonetic tendency both in phonological systems that do not feature PFOD and in those that do. As described in §2, after the first two sound changes in the development of PND (the blurring process) operate, the resulting synchronic alternation PFOD is complex and involves voiced fricatives surfacing in the elsewhere position and voiceless stops in the postnasal position. Example 3 illustrates this alternation.

(3) PFOD: Z → T / N _ , e.g. [**v**oːna] : [ɔm**p**oːna]

At this point, the deviation from articulatory targets described above can cause weak phonetic variation, where voiced fricatives are occasionally produced with occlusion (Ladefoged & Maddieson 1996:137) because occlusion requires less articulatory precision. These articulatory forces result in phonetic variation between voiced fricatives (Z, e.g. [v] or [z]) and voiced stops (D, e.g. [b] or [d]). In other words, the universal phonetic tendency of fricative occlusion causes variation in voiced fricative production across phonological systems and therefore also in those phonological systems that have already undergone the first two sound changes in the blurring process. At some stage, the complex alternation PFOD thus involves voiceless stops (T, e.g. [p] or [t]) in postnasal position and voiced fricatives (Z, e.g. [v] or [z]) that are in phonetic variation with voiced stops (D, e.g. [b] or [d]) elsewhere, due to the universal phonetic tendency toward the occlusion of fricatives. The variation in this latter case is schematized in 4.

(4) PFOD with variation: D ~ Z → T / N _ , e.g. [**b**oːna] ~ [**v**oːna] : [ɔm**p**oːna]

In the initial stages, this variation that crosslinguistically arises from automatic articulatory factors is expected to be highly skewed toward the faithful—in this case, fricative—articulation ([v] or [z]). However, a cognitive preference for simple (albeit unnatural) alternation, which favors the variant with a stop ([b] or [d]), will be confirmed by experimental results in this article. Despite the weakness of the preference for the stop response in the experiment, over time this preference can result in an accelerated reversal of the skewed variation.

For example, a speaker of a language with PFOD is exposed in the majority of inputs to an alternation [**v**oːna] : [ɔm**p**oːna], but occasionally the speaker is also exposed to [**b**oːna] : [ɔm**p**oːna] due to the low-level phonetic process of fricative occlusion. Additionally, they will produce or perceive some of the target fricatives as stops. This is how a regular sound change would operate too: variation based on production and perception can result in a reversal of distribution (from [v] to [b] as the prevalent variant) and, via phonologization, in a sound change. Since this process appears to operate significantly more frequently when it simplifies an alternation, it is reasonable to assume that it is accelerated by a synchronic cognitive mechanism: speakers associate the variant [**b**oːna] with [ɔm**p**oːna] more readily when it is preferred by their grammar. This 'catalyzes' the operation (initiation) of sound change and results in a direct influence of the cognitive bias on typology. In other words, the higher rate and frequency of those sound changes that simplify an alternation result in the observed synchronic typology: alternations resulting from such sound changes accelerated by cognitive bias are more frequent than expected. In fact, because historical bias alone cannot explain the higher frequency of operation of the last sound change in those cases where the sound change simplifies an alternation, I argue that this gradual preference for stop articulation operating on gradient phonetic variation is precisely what catalyzes occlusion of voiced fricatives in the development of PND.

The proposed mechanism, called 'catalysis', can be summarized as in 5.

(5) Catalysis
  a. A subset of segments in an alternation is in passive phonetic variation.
  b. The less frequent variant in this variation is cognitively preferred.
  c. Subjects associate the variant preferred by cognitive factors with a given input more often than the variant that is less preferred (due to learnability, grammatical preferences, or perceptual distance).
  d. Distribution of variation initially skewed toward the faithful variant is reversed toward the variant that is cognitively preferred.

The advantage of catalysis is that it provides a plausible mechanism for how cognitive biases influence phonological typology that has an empirical basis in the present experiment (§4). Catalysis explains the higher rate of the last sound change in the blurring process, which simplifies an alternation and its learning. The mechanism also offers a potential answer to the question of how cognitive factors influence the typology outside of the scope of L1 acquisition (see Bybee 2001 and §1.1)—the proposed mechanism can equally apply during and after L1 acquisition.

In the studies reported in this article, I experimentally simulated conditions for catalysis in the second and third stage of the blurring process that leads to PND and tested it in the artificial grammar learning paradigm. Subjects were trained on two data alternations: PFOD and PND. Subjects were then faced with ambiguous surface forms with voiceless stops postnasally (e.g. [ɔm**p**oːna]) that could go back to a variant with closure consistent with PND ([**b**oːna]) or a variant with a fricative consistent with PFOD ([**v**oːna]). According to catalysis, the first is an innovative variant that arises from low-level phonetic processes (e.g. fricative occlusion) and the second is the faithful variant. Analogous to this lab behavior are speakers in the real world who are faced with phonetic variation resulting from a universal phonetic tendency (e.g. fricative occlusion; see 4). They might generalize the variant with closure more frequently. Over time, this preference due to cognitive bias can result in a higher rate of the reverse distribution of variation (i.e. a higher rate of sound-change initiation), and consequently, this higher rate of operation results in PND being more frequent than is predicted by the historical bias.

When native speakers of two languages with different phoneme frequencies and different voicing realizations are presented with equal amounts of evidence for PND and PFOD, they show a slight preference for the PND response over the PFOD response in the labial series. Subjects thus prefer the simple alternation to the complex one, even if the simple alternation is phonologically unnatural. The link between a synchronic preference for one type of alternation (cognitive bias) and the typological rarity of complex alternations is thus experimentally confirmed, and, crucially, this link cannot be interpreted as part of the historical bias.

4. Experiments. The blurring process assumes a stage with PFOD in the development of PND. Catalysis assumes that speakers of a system with PFOD are faced with PFOD and, occasionally, with variants consistent with PND. Subjects were exposed to singular-plural nonce word pairs with these two alternations: half of the words were consistent with PFOD and half with PND, just as we assume happens in the development of PND.

The advantage of such a design is that the experiments resemble the proposed trajectory of historical changes as closely as possible. The experimental design also follows the approach proposed by Albright and Do (2019), where the same group of subjects is exposed to data consistent with multiple alternations and tested on ambiguous stimuli with both explicit and implicit tasks, which avoids the problem of heterogeneous subject groups in the two conditions and primes implicit rather than explicit learning. The experiment includes evidence for an explicit task, vowel harmony in the feature [±front], and an implicit task, PND and PFOD.

Like any experiment in the artificial language learning paradigm, the experimental design in this article does not completely replicate reality (for a discussion of why this paradigm is nevertheless valid, see Ettlinger et al. 2016). Unlike in the initial stages of catalysis, subjects in the experiment are presented with an equal amount of evidence for PND and PFOD. Such a design was chosen for several reasons.[9] First, it is assumed that in catalysis, initial evidence for PND (which results from low-level phonetic variation) is relatively small. It would be impractical to test synchronic preferences using such small proportions. In laboratory conditions, subjects can be exposed to only a few hundred stimuli. This means that if an experiment contained a small proportion of stimuli containing evidence for PND, subjects would be faced with a handful of PND examples in absolute terms. In the case of actual catalysis, even if the proportion of evidence for PND is small, speakers would be faced with substantially more PND forms in absolute terms. Moreover, in the laboratory, subjects are exposed to each unique item only once, whereas during actual learning, speakers would be faced with the same form several times, thus amplifying the evidence for PND.

The present experiment tests the existence of a synchronic preference for PND over PFOD, all else being equal (i.e. when speakers are presented with equal evidence for the two alternations). If subjects choose disproportionately more PND than is justified by the data, it means there exists a synchronic preference that operates as a pressure every time the speaker needs to make a decision about whether to analyze [ɔmˈpoːna] as a prefixed [ˈvoːna] or [ˈboːna]. There is no specific evidence suggesting that such a preference would operate only when subjects are faced with equal amounts of evidence for PND and PFOD. We can assume that the preference that holds in the least conditioned case (when both are presented equally) also holds in case one variant is more or less frequently represented in the data. One reason to extend this assumption is that, as noted above, in reality speakers would encounter substantially more evidence for the PND variant in absolute terms and would have repeated exposure to individual forms consistent with PND.

Results of the experiments are, however, relevant even without the assumption that the preference for PND operates regardless of the proportions of input data. We saw that catalysis derives typology by explaining why the final sound change operates more frequently. Crucially, fricative occlusion is a required condition for catalysis, regardless of whether the novel stop variant is rare or frequent. It is likely that the preference for PND starts operating when the evidence for the stop-initial variant is very small—the

[9] Future experiments can test preferences given lower proportions of PND-consistent evidence in the training data, but such an experiment will be more challenging to implement and statistically more difficult to model.

synchronic preference for PND can operate every time a speaker needs to make a decision between two variants. But even if the synchronic preference starts operating only when the variation accumulates and reaches a threshold (due to general historical factors) such that the two variants are approximately equally represented (as is the case in the experimental design), the effect would be similar: an accelerated operation and an accelerated completion/phonologization of the final sound change. While in such a scenario the phonetic variation would need to accumulate based on historical factors, the accelerated rate of sound change can still be influenced by catalysis: catalysis predicts not only that the rate of sound change would be accelerated, but also that completion and phonologization would be higher because speakers start analyzing [ɔmˈpoːna] as [ˈboːna] due to the synchronic preference. Without this synchronic preference, the variation would lack the driving force toward its accelerated phonologization and would remain at the rates of the initial phonetic variation or at the rates of sound change from the unconditioned samples.

There are also limitations to the present approach. The experiments only indirectly test whether the observed effects are due to learnability or other synchronic factors: the forced-choice approach in these experiments does not include an incorrect answer. Such a design was chosen for two reasons: not to overburden the subjects (three vs. two choices) and to resemble the assumed catalysis more closely (the assumed trajectory in catalysis involves only two variants—the stop and the fricative variant). Evidence for learning comes from the explicit task on vowel harmony, where subjects choose between correct and incorrect options. While the results suggest that subjects do learn vowel harmony in the explicit task, this does not necessarily entail that learning also occurs in the implicit task (PFOD vs. PND) and that asymmetries in experimental results can be explained by learnability. The preference for PND or PFOD in the implicit task can result from learnability differences (PND is easier to learn), perceptual similarities (e.g. the P-map; Steriade 2001), or general markedness-driven avoidance of voiced fricatives. All three possibilities are relevant to catalysis and are considered part of the cognitive bias, because they operate in individual speakers in a lab setting and likely stem from an association of two variants into a phonological alternation (for historical arguments for this association, see §2). The one confounding factor that would make the results less informative—influence of subjects' native phonologies—is controlled for by conducting experiments in two languages, English and Slovenian, that differ in the frequencies of voiced fricatives and stops.

4.1. Training. Subjects were trained on forming plural nouns from singular nouns in a made-up language called Martian. The plural prefixes were [ɔn-] (before coronals such as [t, d, s] and elsewhere) and [ɔm-] (before labials such as [p, b, f]): for example, singular [ˈsɑnu], plural [**ɔn**ˈsɑnu]. If the singular noun features a front first vowel ([ɛ, i] in English and [eː, iː] in Slovenian), vowel harmony is triggered in the prefix, which then surfaces as [ɛn-] before coronals and [ɛm-] before labials: for example, singular [ˈp^{h}imi], plural [**ɛm**ˈp^{h}imi]. Subjects were explicitly instructed that the plural is formed with <*en-*> and <*on-*> prefixes presented auditorily and orthographically during the instruction phase. Subjects were explicitly asked to pay attention to how plural nouns are formed and were told that the final task would involve forming plural from singular nouns.

The data in the training phase also involved implicit evidence for two alternations: PND (D > T / N _) and PFOD (Z > T / N _). Because the learning of PND and PFOD was tested with an implicit task, experimental instructions never referenced these alternations. Stimuli for the implicit task included items of the shape $C_1V_2C_3V_4$ with an equal number of initial (C_1) labial and coronal voiced stops ([b] and [d]) and initial labial and coronal voiced fricatives ([v] and [z]). In both cases, the plural prefix that ends in a nasal ([ɔn/m-] and [ɛn/m-]) causes the voiced stop to devoice (PND, e.g. [ˈ**b**ɑlu] ~ [ɔmˈ**pʰ**ɑlu]) and the voiced fricative to occlude and devoice (PFOD, e.g. [ˈ**v**ɔnə] ~ [ɔmˈ**pʰ**ɔnə]). Plural forms are thus of the shape {ɛ/ɔ}{n/m}-$C_1V_2C_3V_4$, where C_1 is the 'devoiced' voiceless stop [p] or [t]. In other words, subjects were exposed to the same number of stimuli consistent with PND and with PFOD. In order to prevent subjects from analyzing PFOD as a simple restriction on postnasal fricatives, subjects were also given evidence that voiceless stops ([p] and [t]) and voiceless fricatives ([f] and [s]) remain unchanged in postnasal position ([mp], [nt], [mf], and [ns], respectively). Table 5 schematically represents the alternations in the training phase, and Table 6 lists some actual examples of the stimuli. One disadvantage of the present experiment is that L1 phonology can affect the complexity of the tested alternations. Both English and Slovenian feature voiced stops in the elsewhere condition, but at the PFOD stage, voiced stops are absent from the system. This would mean that [±voice] does not change automatically in the experiment, but it does so in the development of PND. We can still assume that changing two features is more complex than changing one (even if one changes automatically). Under the perceptual distance hypothesis, this aspect is of course not problematic.[10] For all details of the experimental design and structure of the stimuli, see §1 of the online supplementary materials.[11]

	[−voice]		[+voice]	
	[−cont]	[+cont]	[−cont]	[+cont]
# _	T	S	D	Z
N _	T	S	T	T

TABLE 5. Alternations in voiceless and voiced stops and fricatives according to position (implicit task).

	[−voice]				[+voice]			
	[−cont]		[+cont]		[−cont]		[+cont]	
	[LAB]	[COR]	[LAB]	[COR]	[LAB]	[COR]	[LAB]	[COR]
SINGULAR (# _)	[ˈ**pʰ**ɔrɔ]	[ˈ**tʰ**ɑru]	[ˈ**f**urə]	[ˈ**s**ɑnu]	[ˈ**b**ɑlu]	[ˈ**d**ɔru]	[ˈ**v**ɔnə]	[ˈ**z**ɔlɛ]
PLURAL (N _)	[ɔmˈ**pʰ**ɔrɔ]	[ɔnˈ**tʰ**ɑru]	[ɔmˈ**f**urə]	[ɔnˈ**s**ɑnu]	[ɔmˈ**pʰ**ɑlu]	[ɔnˈ**tʰ**ɔru]	[ɔmˈ**pʰ**ɔnə]	[ɔnˈ**tʰ**ɔlɛ]

TABLE 6. Examples of Martian words for the implicit task in the English experiment.

During the training phase, the stimuli (as described above and summarized in supplementary materials §1) were presented in a randomized order (randomized for each subject). A unique picture of a Martian creature was associated with each stimulus (also randomized for each subject).[12] Orthography was generally not given with the training stimuli, but to prompt subjects to focus on the experiment, ten randomly chosen words (randomized for each subject) in the training phase were presented orthographically as well as auditorily in the singular. The plural form was then given only in audio form, and subjects had to enter a transcription of the form they heard. This orthographic task was never used for words that began with voiced stops or fricatives (i.e. words bearing

[10] The data also involved place assimilation of the prefix nasal (e.g. [n] if coronal, [m] if labial). This assimilation was never tested experimentally; its primary purpose was articulatory ease during stimuli recording and achieving conformity with subjects' L1 phonologies, such that no attention was attracted to the distribution of nasals.

[11] The supplementary materials referenced throughout can be accessed at http://muse.jhu.edu/resolve/143.

[12] Pictures of Martian creatures in the experiments were taken from van de Vijver & Baer-Henney 2014 with the authors' permission.

evidence for PND and PFOD) in order to minimize the effect of orthography and of increased memorization due to this task.

While PFOD is not an unnatural alternation according to the definition in §2, in the sense that postnasal frication and voicing (the opposite process of PFOD) are not universal phonetic tendencies, at the same time, it is unnatural in the sense that when PFOD applies, stops surface as voiceless in postnasal position, where they are universally dispreferred (see Beguš 2019 for extensive discussion). Unlike PND, however, PFOD allows subjects to construct alternative grammars that are not necessarily unnatural, especially if the two processes are treated separately in phonology. If no evidence existed that voiceless fricatives remain unchanged postnasally and voiced stops devoice postnasally, PFOD could be analyzed as a restriction against postnasal continuants and a global restriction against voiced stops (*[+nasal][+cont] and *[+voice, −cont] markedness constraints in OPTIMALITY-THEORETIC terms; Prince & Smolensky 2004 [1993]). This is exactly the synchronic system of stage 2 in the development of PND. Even if voiceless fricatives remain unchanged, PFOD can be analyzed as resulting from *[+nasal][+cont] and *[+voice, +cont] constraints, with an additional higher-ranked faithfulness constraint that preserves the identity of voiceless fricatives, although such an analysis is less likely. Because the experimental design requires evidence for voiced stops to remain unchanged except postnasally (as is also predicted by catalysis), it is possible that subjects analyze PFOD as complex and unnatural. Regardless of whether subjects treat PFOD as an isolated, complex, and motivated process or as unnatural and complex, the results have direct implications for modeling synchronic influences on typology (see §6).

4.2. TEST. Learnability of the two alternations was tested with an implicit task because implicit learning might resemble L1 phonological acquisition more closely than explicit learning does (on this question, see Moreton & Pertsova 2017 and Moreton et al. 2017, but further research is warranted). Subjects were not given any instructions on the implicit task. The alternations in the implicit task are summarized in Tables 5 and 6 above, and in Figure 3.

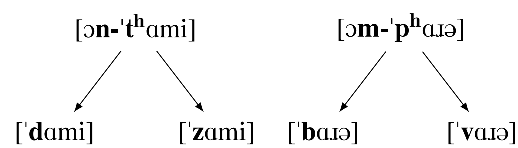

FIGURE 3. Examples of ambiguous stimuli in plural form with a prefix and the corresponding possible singular forms subjects were asked to choose from.

After the training period, subjects were told that they would hear some Martian words they had not heard before and that they would be asked to indicate the most likely way to say new words in Martian. In the test phase, subjects were given ambiguous stimuli in the plural form of the shape {ε/ɔ}{n/m}-$C_1V_2C_3V_4$, with a devoiced noun-initial stop (C_1 = [p] or [t]) that could go back to either a voiced stop [b] or [d] (consistent with PND) or a voiced fricative [v] or [z] (consistent with PFOD). Subjects had to choose between two singular forms: one with a word-initial voiced labial or coronal fricative ($\mathbf{Z}V_2C_3V_4$, consistent with PFOD), and one with a word-initial voiced labial or coronal stop ($\mathbf{D}V_2C_3V_4$, consistent with PND). Figure 3 exemplifies this task.

Preference for one option over the other would suggest either that (i) one alternation is easier to learn than the other, (ii) subjects choose the option that is typologically more frequent and/or articulatorily easier and perceptually more salient (less marked, in phonological terms), or (iii) subjects use knowledge of phonemic distributions in their native language and match the phoneme frequencies of their native language to their experimental responses. Both (i) and (ii) are informative for our purposes (see discussion in §4.5 and §5.2 on how we control for (iii)).

4.3. PROCEDURE. The test phase consisted of the two tasks, one explicit and one implicit. The order of items testing the explicit and implicit tasks was mixed and randomized for each subject. In the explicit task, the experiment tested whether subjects learned the vowel harmony in plural prefixes. Six items in the singular were created for this task, three that trigger frontness harmony and three that do not: [rema], [liro], [leni] vs. [lonu], [ruro], and [lona]. Subjects were presented with the six items in the singular (both orthographically and auditorily), each paired with a picture of a Martian creature (see Figure 4). After they heard the item, subjects were shown four pictures of the same creature and were given two orthographic stimuli to choose from: one showing the correct harmony, and one showing the incorrect disharmony. For three of these items, the correct response appeared on the left button, and for three on the right button. The explicit task included orthographic choices in order to avoid overburdening of subjects with auditory stimuli: the auditory-only stimuli are reserved for the more informative implicit task, which never included orthography. Button-side ordering was kept constant for all subjects and items, but the ordering of individual items was randomized.

For the implicit task, subjects were first presented with four identical pictures of a Martian creature accompanied by a recording of the test word in the plural without any orthographic stimuli. The button that played the recording was embedded in a written sentence 'These are [▷]'. After playing the sound, subjects were presented with a single picture of the same Martian creature. The subjects played the first stimulus, presented in a sentence 'Is this a [▷],'; the second stimulus then appeared, embedded in 'or a [▷]?'. After the subjects heard both stimuli, they were asked 'Which one is it?'[13] and were given two choices ('1st' and '2nd'). Stop-initial and fricative-initial singular forms were equally assigned the first or second position, so that the position of the answer would not influence the results. In other words, the stop-initial response appeared as the first option in half of the eight items, and as the second option in the other half. The ordering was kept constant for all subjects and items, but ordering of individual items was randomized. That stimuli ordering does not affect the results is suggested by the statistical analysis given in §5. None of the stimuli in the implicit task were presented orthographically.

The stimuli for the implicit task consisted of eight plural items of the shape *prefix*-$C_1V_2C_3V_4$, where the prefix was [ɔn, ɛn, ɔm, ɛm] with the correct harmony based on V_2 (for an example, see Fig. 3). C_1 was a voiceless ('devoiced') labial stop in four items and a voiceless ('devoiced') coronal stop in four items. Four items included a noun with a front first vowel (V_2) and four with a nonfront first vowel (two for each place-of-articulation group). C_3 consisted of [m, n, r, l, j, w]. Vowels V_2 and V_4 consisted of [ɑ, ɛ, ɔ, i]. Test words were created such that there were no minimal pairs with the training words. Each plural form was matched with two corresponding singular forms of the shape $C_1V_2C_3V_4$, one with a voiced stop in C_1 and one with a voiced fricative. A sample spectrogram of a stimulus [ɔmpara] and the two possible responses [bara] and [vara] are given in the supplementary materials (figure 1). Figure 4 illustrates the full experimental interface.

4.4. SLOVENIAN. The Slovenian experiment was identical to the English experiment, except as described in this section and in supplementary materials §1.2. Unlike the En-

[13] The question was followed by '(you can replay all three words)'.

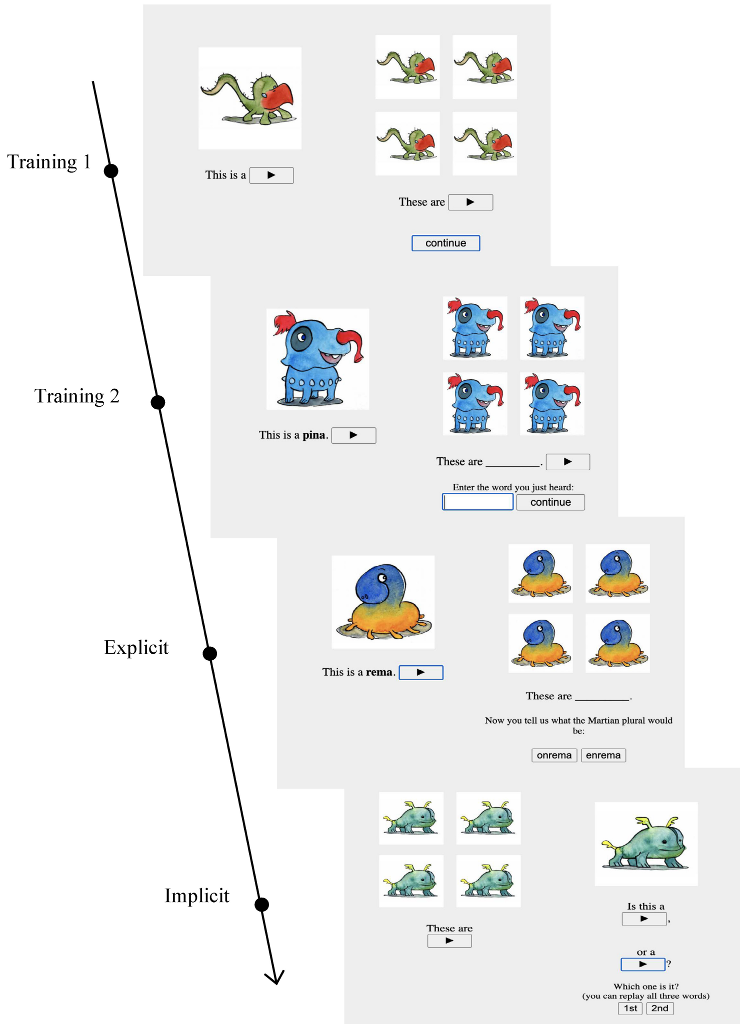

FIGURE 4. Examples from the Experigen (Becker & Levine 2013) interface as used in the experiment (converted from color to grayscale in the print version of this article) for the training phase and test phase. Pictures of Martian creatures in the experiments are from van de Vijver & Baer-Henney 2014, used with permission.

glish experiment, which was conducted entirely online with subjects recruited via Amazon MTurk, the Slovenian experiment was conducted in person in Slovenia, with the supervision of research assistants. Subjects were native speakers of Slovenian, a south Slavic language with approximately 2.2 million speakers (Simons & Fennig 2018). Subjects were recruited from the general public with the help of research assistants (via personal contacts and social media). Altogether 150 subjects participated in the Slovenian experiment.

Test items for the explicit task included six $C_1V_2C_3V_4$ words with C_1 = [l, r] and three front V_2s (the other three were nonfront). Subjects were asked to choose between the correct harmonic and incorrect disharmonic forms with the prefixes [ɔn/m-] and [ɛn/m-]. Subjects were tested on the implicit task with twelve trials. First, subjects were presented with a plural form of the shape *prefix*-$C_1V_2C_3C_4$, where C_1 is a devoiced [p] or [t] that can go back to a voiced stop [b, d], consistent with PND, or to a voiced fricative [v, z], consistent with PFOD. The 'devoiced' stop is a labial in half (six) of these trials, and a coronal in the other half, In half of the items, V_2 is nonfront (therefore with prefix [ɔn/m-]); in the other half, V_2 is front (therefore with prefix [ɛn/m-]). All experimental stimuli are given in the supplementary materials §1. After being exposed to the plural forms, subjects were asked to choose between a corresponding singular form with an initial voiced stop and one with a voiced fricative. Both options were presented to the subjects in each trial, and were given in auditory form only, without orthographic inputs. All other properties of the test phase are the same as described for the English experiment in supplementary materials §1.1.1.

4.5. Why two languages. The possibility that a preference for one or the other response is due to the phonemic frequencies of subjects' L1 phonologies would make this task uninformative, which is precisely why the experiment was conducted in two languages, English and Slovenian. In both, coronal /d/ is more frequent than /z/, but in Slovenian the labial fricative /v/ is more frequent than the labial stop /b/, and in English the distribution is the opposite, or the difference is at least substantially smaller than in Slovenian. The difference in frequency between /b/ and /v/ in Slovenian is substantial, persists in word-initial position, and is established based on several lemmatized and nonlemmatized corpora of both written and spoken Slovenian (Suhadolc 2013, Marvin et al. 2018). Word-initially, /b/ has a relative frequency of 3.6%, while the relative frequency of /v/ is 5.3% (see Suhadolc 2013). In other corpora, the difference across positions is between 1.7–1.9% and 3.3–4.2% (see Marvin et al. 2018).[14] For English, by contrast, different studies report slightly different results for the relative frequencies of /b/ and /v/. In most studies or corpora, /b/ is more frequent than /v/ (Wang & Crawford 1960, Mines et al. 1978, Kessler & Treiman 1997). This is true across positions as well as in onset position (Kessler & Treiman 1997). One study shows that the distribution of the frequencies of /b/ and /v/ is the opposite (Hayden 1950; also in some corpora by Wang and Crawford (1960)), but the differences here are minor. If subjects simply matched native phoneme frequencies in their experimental responses, we would expect a higher response rate for /v/-initial words in Slovenian and for /b/-initial words in English. For an even stronger piece of evidence, based on bigram frequencies, that the L1 phonologies do not crucially affect the results, see §5.2.

English and Slovenian are also suitable for testing the learning of alternations involving the feature [±voice] because phonetic realization of the phonological contrast differs substantially in the two languages. In fact, many analyses assume the phonological feature involved in the English contrast is [±spread glottis] rather than [±voice] (Iverson & Salmons 1995 and the literature therein). English voiceless stops are realized with a substantial period of aspiration, especially in the onset position of stressed syllables. Voiced stops are partially or fully devoiced, depending on the position. Utterance-initially, voicing is often lacking completely (Davidson 2016). Slovenian, by contrast, is a 'true voicing' language in which voiced stops are fully voiced in all positions, including utterance-initially (with prevoicing; see supplementary materials figure 2 and Toporišič 2004). The two languages also differ in the exact realization of /v/. In English,

[14] Slovenian phoneme frequencies are calculated based on grapheme frequency in Suhadolc 2013, but because there is a strong tendency toward one-to-one correspondence of the Slovenian consonant inventory and orthographic representation (especially of /b/ and /v/ in initial position), the results of the analysis based on graphemes can be extended to phonemes. Marvin et al. (2018) account for the grapheme-phoneme differences.

labiodental /v/ is always analyzed as a fricative, while in Slovenian descriptions vary between a fricative /v/ (Toporišič 2004) and an approximant /ʋ/ (Šuštaršič et al. 1995). The exact acoustic distinction between the two is, however, challenging to establish. The nature of the bilabial fricative/approximant and its distribution, which is potentially dialectal in Slovenian, warrants a separate discussion, but our recordings indicate the presence of frication noise (an example is given in supplementary materials figure 3).

In addition to controlling for the voicing realization and phoneme frequencies, the experiments were designed with the goal of diversifying the subject pool, keeping the ratio of stimulus vs. test items as high as possible, and maximally balancing experimental stimuli. For these reasons, speakers of the two languages tested were recruited from the general public. The experiment with English speakers was conducted online; the Slovenian experiment was conducted in person with the supervision of research assistants and with controlled experimental equipment. The ratio of training to test items is 32:8 for the explicit task in English and 32:12 in Slovenian, in order to prevent the test phase itself from influencing the responses. The balancing of experimental stimuli is described in detail in supplementary materials §§1.1.1 and 1.2.1.

4.6. Subjects. A total of 353 subjects participated in the two experiments. Of the 203 English participants, 198 finished the experiment in full and completed a final demographic questionnaire. Of the 150 Slovenian participants, 141 completed the test phase and finished the final questionnaire. Subjects who indicated that either English or Slovenian, respectively, was not their native language or who indicated that they had linguistic education (or had taken any classes in linguistics as part of their education) were excluded from the analysis (subjects with no responses on the two questions were included). Altogether, 170 subjects in the English group and 110 in the Slovenian group were analyzed on the implicit task. The 170 subjects in the English experiment provided 1,346 analyzed responses on the implicit task; the 110 subjects in the Slovenian experiment provided 1,317 responses. A total of 280 analyzed subjects is a relatively high number of participants compared to most similar artificial grammar learning experiments, especially for the in-person experiment. For other details on the subjects and exclusion criteria, see supplementary materials §1.4.

5. Results. Responses for the explicit (correct vs. incorrect) and implicit (PND vs. PFOD) tasks across the two experiments (English and Slovenian) and across the two places of articulation and vowel frontness are given in Tables 7 and 8. Raw counts reveal that the correct response (consistent with vowel harmony) was more frequent than the incorrect response for all groups in the explicit task.

	English			Slovenian		
	correct	incorrect	% correct	correct	incorrect	% correct
front	313	193	61.9%	226	104	68.5%
back	311	196	61.3%	207	122	62.9%

Table 7. Raw counts of the number of correct (harmonic) and incorrect (disharmonic) responses on the explicit task across the two languages and vowel frontness.

	English			Slovenian		
	stop	fricative	% stop	stop	fricative	% stop
labial	388	286	57.6%	397	261	60.3%
coronal	354	319	52.6%	341	319	51.7%

Table 8. Raw counts of the number of stop and fricative responses on the implicit task across the two languages and places of articulation.

Likewise, the raw counts suggest that the response consistent with PND (stop response) was more frequent than the response consistent with PFOD (fricative response) for all four groups on the implicit task. This effect appears to be substantially stronger for the labial series of stops than for coronals. The preference for the correct response, consistent with vowel harmony, seems more robust than the preference for the stop response (PND), but the difference is not very substantial, especially given that the evidence for the first is explicit and categorical (occurring in every one of the fifty-eight to sixty items) without any ambiguous stimuli. By contrast, the preference for the PND-consistent response emerges from the thirty-two items in the training phase that bear an equal amount of evidence for both PND and PFOD: devoiced postnasal stops can go back to either voiced stops or voiced fricatives. Additionally, the PND-consistent response emerges in an implicit task without any direct instructions to the subject.

Figure 5 plots counts of subjects according to how they performed on the implicit task (number of responses consistent with PND). Subjects who always selected the fricative (PFOD) response have a score of 0. Subjects with all stop responses (PND) have a score of 8 or 12, respectively (English vs. Slovenian). For reference, the plots also show predicted values if subjects responded randomly, indicated by vertical lines with dots. These predicted values are calculated from a binomial distribution with n = 8 or 12, respectively, $p = 0.5$. For each k, $\Pr(k; n, p)$ was calculated from a binomial distribution and multiplied by the number of subjects in each sample: 167 and 109, respectively.

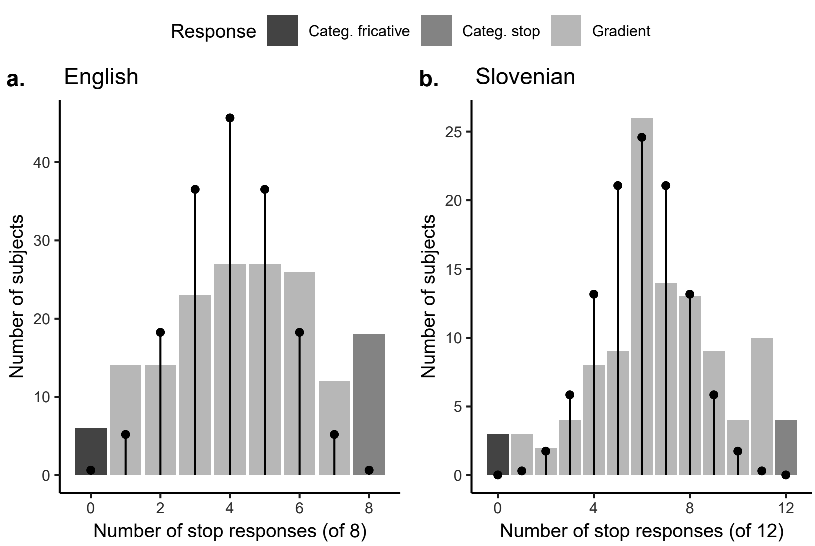

FIGURE 5. Raw counts of subjects according to the number of stop responses in the (a) English and (b) Slovenian experiment.

The distribution of subjects is substantially higher than expected on the two marginal ends of the stop-fricative opposition and is lower than expected in its middle (in the English experiment). There are more subjects who chose the stop response more often than subjects who chose the fricative response more often (especially in the labial series). Especially in the English experiment, categorical responders who chose the PND-consistent response are notably more numerous.

5.1. EXPLICIT TASK. To confirm that the subjects were learning the explicit alternation (vowel harmony), the results of the explicit task were fit to two logistic regression mixed-effects linear models, with harmonic (correct) responses as successes and disharmonic (incorrect) responses as failures, one for each experiment. For an exact description of the model, see supplementary materials §1.5. Correct answers (consistent with the explicit vowel harmony) are significantly above chance level in the English experiment ($\beta = 0.51$, $z = 5.6$, $p < 0.0001$) with 95% profile CIs at [57.8%, 67.2%]. The Slovenian experiment yields similar results. The harmonic responses that conform to the training data are significantly more frequent than the disharmonic responses ($\beta = 0.70$, $z = 6.72$, $p < 0.0001$) with profile CIs at [61.9%, 71.9%].

The results suggest that subjects learned the explicit alternation above chance. Similar outcomes in both experiments suggest that English-speaking subjects, who were recruited via Amazon MTurk, were not supervised during the experiment, and were not given the same headphones and the DAC amplifier as in the Slovenian experiment (see supplementary materials §1.2.1), did not perform substantially worse on the explicit task compared to the in-person experiment participants. On average, the in-person Slovenian group performed slightly better than the English Amazon MTurk group.

5.2. Implicit task. To test the significance of the preference for the stop response (consistent with PND) vs. the fricative response (consistent with PFOD), the data were fit to two logistic regression mixed-effects linear models, one for each experiment. Coefficients of the final models are given in Table 9; estimates with confidence intervals across places of articulation and across the two languages are given in Figure 6. The significance of the predictors of interest is the same in the full models, including all interactions and random slopes, and in the final chosen models; here I report the nonfull models because of interpretability and to signal which predictors are significant. For a detailed description of the models, see supplementary materials §1.6.

		ENGLISH			
		EST	*SE*	*z*-value	Pr(>|*z*|)
(intercept) =	labial	0.43	0.21	2.02	0.04
	coronal	−0.29	0.28	−1.03	0.30
		SLOVENIAN			
		EST	*SE*	*z*-value	Pr(>|*z*|)
(intercept) =	labial	0.62	0.22	2.78	0.01
	coronal	−0.56	0.32	−1.79	0.07
	HARMONY	−0.20	0.13	−1.49	0.14

Table 9. Coefficients of the final model (English and Slovenian).

Subjects chose the stop response (compared to the fricative response) in the English experiment significantly more frequently for the labial series of stops ($\beta = 0.43$, $z = 2.02$, $p = 0.04$).[15] In the coronal series, no such difference is observed (based on confidence intervals; see Fig. 6). Given that subjects were recruited via Amazon MTurk in the English experiment and some potentially may not have paid attention during the experiment, it is reasonable to also test responses in the data from which the subjects who scored the lowest on the explicit task are removed. If we include only subjects who scored 50% or better on the explicit task, the preference for the fricative response remains significant ($\beta = 0.58$, $z = 3.0$, $p = 0.003$).

The same procedure is used to test significance in the Slovenian experiment. For a detailed description of the model, see supplementary materials §1.6. The estimates are given in Table 9. Subjects prefer the stop response in the labial series ($\beta = 0.62$, $z = 2.78$,

[15] Profile CIs cross the zero in the model not corrected for underdispersion.

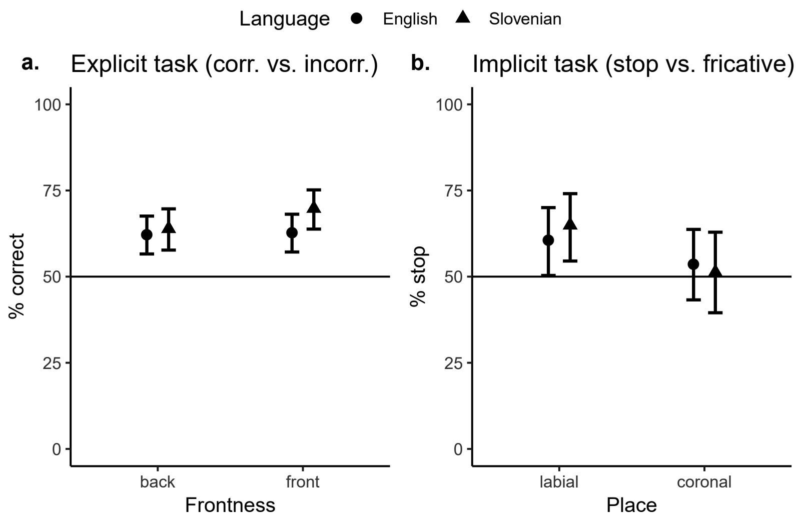

FIGURE 6. (a) Logistic regression mixed-effects model with correct responses on the explicit task as the dependent variable and frontness as the predictor, with interactions (with 95% CI), for each of the two languages tested. (b) Logistic regression mixed-effects model with percentage of stop responses on the implicit task as the dependent variable and place as the predictor (in the Slovenian model, harmony is also a predictor) (with 95% CI).

$p = 0.005$), but again no such effect is observed in the coronal series. The same effect persists if we exclude nonlearners (those who scored less than 50% on the explicit task): $\beta = 0.54$, $z = 2.5$, $p = 0.01$.

5.3. INFLUENCE OF L1 PHONOLOGY: BIGRAMS IN SLOVENIAN. A strong piece of evidence against the possibility that subjects simply matched frequencies from their L1 phonologies comes from bigram frequencies and absolute response rates in the experiments. Frequencies of the tested word-initial bigrams in lemmas (from the standard dictionary of Slovenian; Bajec et al. 2000) are not distributed equally. /b/ is more frequent than /v/ before /a/ (688 vs. 341 in counts) and before /u/ (267 vs. 32). /v/ is more frequent than /b/ before /i/ (476 vs. 388) and /e/[16] (782 vs. 489). To test nonlemmatized frequencies in a spoken corpus, word-initial sequences of /b/ and /v/ followed by the vowels /a/, /i/, /e/, and /u/ were extracted from GOS (Korpus govorjene slovenščine; Verdonik & Zwitter Vitez 2011), a nonlemmatized transcribed corpus of spoken Slovenian. Word-initial /v/ is more frequent than /b/ before /a/ (3,542 vs. 1,353) and before /e/ (15,779 vs. 1,852), but /b/ is more frequent than /v/ before /i/ (13,074 vs. 5,059) and /u/ (693 vs. 221).

Learnability of PND and PFOD is tested on items with initial /b-/ and /v-/ before six vowels: /i/ (twice), /e/, /a/ (twice), and /u/. The /b/-initial response is more frequent than the /v/-initial response in five of six items in the experiment. /b/ is a more frequent response before items with V_2 /a/ (two items), /e/, /i/, and /u/. In only one item with V_2 /i/ is the /vi/-initial response more frequent than the /bi/-initial response (being given fifty-six vs. fifty-three times). In other words, subjects prefer the /b/-initial response even in those stimuli in which the /v/-initial bigram in the native phonology is more frequent than the /b/-initial bigram, estimated from a lemmatized dictionary and nonlemmatized

[16] In this paragraph, /e/ stands both for /eː/ and /ɛ/, because the corpora do not distinguish them.

spoken corpus. This means that phoneme frequency distribution in the native language likely does not influence the experimental outcomes. To my knowledge, this experiment is one of the first to test learnability of an alternation on two languages with different relative frequencies of the tested phonemes.

6. Discussion. When trained on data with an equal amount of evidence for PND and PFOD and tested on the ambiguous stimuli, subjects show a small but significant preference for the stop response (consistent with PND) compared to the fricative response (consistent with PFOD) in the labial series. These results are not likely to be influenced by subjects' native phonological grammars, because English and Slovenian differ in phonemic frequencies of /b/ and /v/, in how their bigram frequencies are distributed, and in the phonetic realization of the feature [±voice] (or [±spread glottis]). Preference for the stop response (PND) appears to be stronger in the Slovenian than in the English experiment; if phoneme frequencies affected the results, an opposite distribution would be expected (/v/ is more frequent than /b/ in Slovenian). Moreover, because the learning of the two alternations is tested with an implicit task, it is likely that the results more closely resemble L1 phonological acquisition (compared to the explicit task) (Moreton & Pertsova 2017, Moreton et al. 2017). Additionally, recruitment for the experiment was done on the general public without prior linguistic experience in two experimental modes: online and in person.

While the preference for the PND response appears weak, comparing the results of the implicit task, in which both options are correct (and presented with equal frequencies in the training data), to the results of the explicit task, in which there is a correct and an incorrect option, reveals that the preference for a stop-initial response in absolute terms is comparable to the preference for the correct response. Figure 6 illustrates that the proportion of correct responses on the vowel harmony pattern is comparable to the preference for the stop-initial response.

The advantage of the proposed model of combining experimental data with statistical modeling of sound change is that the link between experimental responses and the typological distribution of phonological alternations cannot be attributed to historical bias. As discussed in §2.3, the historical-bias approach alone does not derive the observed typological distributions: PFOD is less frequent than predicted by the historical-bias approach, whereas PND is more frequent than expected. Additionally, the last sound change that leads to PND operates significantly more frequently when it simplifies the resulting alternation and consequently simplifies its learning. These nonexperimental observations already suggest that cognitive biases directly influence the observed typology. Experimental data in this article suggest that PND is favored over PFOD by the cognitive factors tested in the experiment. In other words, the typological distribution that cannot be explained within the historical-bias approach matches the experimental responses. The experimental results thus confirm the link between cognitive bias and typology. Experimental results suggest that the main manifestation of the cognitive factor is avoidance of complex alternations (in line with previous work; see Moreton & Pater 2012a,b) and that we can experimentally confirm this link even when the historical factor is controlled for. Additionally, the experiments simulate a historically plausible scenario for how a synchronic preference for the stop response can directly result in the acceleration of a sound change operation, which consequently results in typological distributions (the catalysis model).

Results also suggest that the preference for the PND response ([b] vs. [v] and [d] vs. [z]) is stronger for labials than for coronals. This asymmetry in experimental results is not undesirable, however. A typological survey of PFOD and PND in Beguš 2019 reveals that the last sound change, occlusion of fricatives to stops, often does not include all places of articulation, but targets only a subset. It is thus possible that the asymmetry in the experimental results reflects the actual asymmetry in the observed typology. In other words, the experimental results on learnability are consistent with the typology, where only a subset of places of articulation are targeted.

There are at least two possible explanations within the cognitive-bias approach for why subjects prefer the response consistent with PND. Both possibilities are informative for modeling historical and cognitive influences on typology. First, it is possible that one alternation is easier to learn than the other. This possibility is consistent with a large body of research showing that complex alternations are more difficult to learn than simple alternations (i.e. complexity bias; see Moreton & Pater 2012a,b). Additionally, the higher-than-chance response in favor of the correct harmonic option (see §5.1) suggests that subjects indeed learned the alternation in the explicit task. It is possible that a similar learning mechanism underlies the preference for the PND response on the implicit task. The novel aspect of the present experiment (besides controlling for historical bias) is that the preference for simple alternation emerges even if the simple alternation is unnatural, that is, operating in exactly the opposite direction from universal phonetic tendencies. To my knowledge, in none of the experiments testing complexity bias thus far has the alternative alternation been unnatural.

The second possibility is that subjects prefer the variant that is articulatorily or perceptually less marked: either the variant that requires less articulatory effort when presented with ambiguous data or the variant that is perceptually closest to the stimulus. Voiced fricatives (such as [v, z]) are articulatorily dispreferred (Ohala 1983, 1997, 2006:688, Ladefoged & Maddieson 1996:137, Smith 1997) and typologically less common compared to voiced stops (such as [b, d]) (Moran et al. 2014). Such dispreferred segments have been labeled ‘marked’ in phonological theory (de Lacy 2006a).[17] It is possible that subjects use tacit universal phonological knowledge and choose the less-marked option or the variant that requires less articulatory effort.

This line of explanation faces some challenges. First, while voiced fricatives are generally more marked than voiced stops, it is unclear whether they are more marked for L1 speakers of languages in which voiced fricatives are more frequent than voiced stops. As discussed in §4.2, voiced labial fricatives are more frequent than voiced labial stops in Slovenian, and some of the tested bigrams are more frequent if they involve an initial voiced fricative. According to the markedness hypothesis, subjects would have to disregard phonemic frequency and choose their responses based only on the general markedness or articulatory effort. Moreover, if markedness avoidance is the main mechanism behind subjects’ experimental responses and if experimental responses are indicative of what happens in phonological development, we would expect the same rate of markedness avoidance and consequently the same rate of sound change occlusion of fricatives (/v/ > [b]) in all cases, regardless of whether the sound change simplifies an alternation (and therefore its learnability; see §2.3). Because the rate is not equal in the two conditions, it is unlikely that articulatory markedness affects the results. Even if articulatory markedness affects experimental outcomes, the results are nevertheless informative for our purposes. The fact that a higher rate of sound change occurs only when two segments are connected with a synchronic alternation provides evidence in

[17] See de Lacy 2006a for a discussion of markedness and its diverse definitions.

favor of the link between the experimental results and typology (and therefore supports the catalysis model).

One of the more challenging questions in phonology is whether experimental evidence of the complexity bias is influenced by featural or perceptual complexity. It has been assumed that the more features an alternation manipulates, the more complex it is and therefore more difficult to learn. Featural complexity is, however, highly conflated with perceptual complexity. It is thus possible that the PND alternation is preferred and easier to learn compared to the PFOD alternation because the perceptual distance between T and D ([p] vs. [b] or [t] vs. [d]) is smaller than that between T and Z ([p] vs. [v] or [t] vs. [z]) (in line with the P-map hypothesis: Steriade 2001; see also Wilson 2006, White & Sundara 2014, White 2017).

To quantitatively evaluate the role of perceptual distance in the synchronic preference in the experimental results and potentially in the process called catalysis, I estimated *d′* values for the perceptual distance between the voiceless labial and coronal stop and corresponding voiced stops and voiced fricatives. Consonant confusion matrices were taken from listeners of four languages (including English) from Singh & Black 1966. The data were fit to a bias-reduced probit regression linear model coded such that it estimates *d′* values and performs significance testing on differences in *d′* values (for a detailed description of the data and models, see supplementary materials §2). [p] and [b] are significantly more similar perceptually than [t] and [d] are (the difference between *d′* of [p] ~ [b] and [t] ~ [d] is estimated at $\beta(\Delta d') = 0.84$, $z = 2.17$, $p = 0.03$). Certainly, the difference in *d′* between [p] ~ [b] and [p] ~ [v] is significant: $\beta(\Delta d') = 1.45$, $z = 3.46$, $p = 0.0005$. In other words, the perceptual distance between a voiceless and voiced labial stop ([p] ~ [b]) is smaller than that between a voiceless stop and voiced fricative ([p] ~ [v]); the perceptual distance between a voiceless and voiced labial stop ([p] ~ [b]) is smaller than that between a voiceless and voiced coronal stop ([t] ~ [d]). This could explain the synchronic preference for stop responses in the experiments: subjects prefer perceptually minimal alternations. The P-map hypothesis (Steriade 2001) similarly claims that the preference for perceptually minimal alternations is part of the synchronic grammar. It is possible that the synchronic preference in the experiment is strongest in the labial series precisely because [p] and [b] are perceptually most similar.

Conclusive tests of the two hypotheses (featural vs. perceptual complexity) are difficult to design. The results of the experiment presented here remain relevant even if perceptual complexity plays a role in experimental responses (see §3). Under this approach, the causal factor behind the typology-perception link still needs to be tied to subjects' internal preference for the similarity of segments involved in an alternation. In other words, while perception plays a role both in cognitive- and historical-bias approaches (§1), it is likely the case that cognitive factors—relationship to a synchronic phonological alternation—accelerate the operation of sound change and consequently result in typology. If only perceptual difference, regardless of its association with a phonological alternation, influenced the typology, the same rate of application of fricative occlusion would be expected both where the sound change simplifies an alternation and where it does not (see §§2 and 3). Regardless of its underlying driving forces, the synchronic preference for the stop (simple) response has been confirmed in two experiments on two languages with 280 subjects in at least a subset of places of articulation. The observed preference is reflected in the typology and, crucially, is likely not influenced by historical-bias factors.

7. CONCLUSIONS AND FUTURE WORK. Phonology offers a unique test case in the discussion of historical and cognitive influences on human behavior. Combining statistical modeling (Beguš 2020) and experimental work, this article presents a framework for testing one approach while controlling for the other. The crucial aspect of this framework is identifying mismatched predictions between the cognitive and historical approaches and experimentally testing one approach while controlling for the other. The results of one set of such experiments presented here suggest that when subjects are trained on PFOD and PND and tested on ambiguous data, they show a weak but significant preference for the PND response. The results point to a causal link between cognitive bias and typology: the PND alternation that is synchronically preferred in a laboratory setting is also typologically more frequent than is predicted by historical bias alone. Crucially, the link between cognitive bias and typology does not have a competing historical explanation (unlike in many other experiments), because the historical-bias approach makes the opposite prediction in these types of cases.

The results thus yield insights into those aspects of phonology that are primarily influenced by human cognition and are not emergent from language's transmission. The data suggest that the cognitive part of phonology is responsible for avoiding (featurally or perceptually) complex alternations and keeping phonology structurally simple. Applying the framework to further alternations should provide a better understanding of which aspects of phonological grammar and typology are emergent from historical factors and which aspects are primarily influenced by cognition.

Some of the features in the experimental design that control for various nonlinguistic variables are novel to the paradigm. The results are likely not influenced by subjects' L1 phonologies since Slovenian and English have different distributions of relative frequencies of both the tested phonemes and the tested word-initial bigrams. The experimental design aimed to diversify the subject pool by recruiting from the general public, both online and in person with controlled and uncontrolled auditory presentation conditions and excluding subjects with prior linguistic experience.

Finally, the presented experimental results combined with statistical models of sound change point to one potential mechanism for how cognitive bias directly influences the typology—catalysis. Universal phonetic tendencies operate crosslinguistically and cause variation in surface forms. Initially, the variation is heavily skewed toward the original, faithful variant. If the nonfaithful variant simplifies learning, however, learnability preferences can skew the initial faithful distribution into a system that favors the featurally or perceptually less complex variant. This is precisely what likely underlies the higher rate of operation (initiation and phonologization) of fricative occlusion in the development of PND. Catalysis has several advantages—it explains the higher rate of operation of those sound changes that simplify an alternation, is able to derive sound change after the L1 acquisition period, and provides an empirically based explanation for a question that is inherently difficult to answer—how biases in human cognition in general result in the observed distribution of patterns in human language. Exploring further implications of the proposed mechanism should yield a better understanding of the relationship between the historical transmission of language and the role that cognition plays in the phonological system of human language.

REFERENCES

ALBRIGHT, ADAM, and YOUNGAH DO. 2019. Three biases in learning phonological alternation. Cambridge, MA: MIT, and Hong Kong: University of Hong Kong, MS.

Altarriba, Jeanette. 1993. The influence of culture on cognitive processes. *Cognition and culture: A cross-cultural approach to cognitive psychology* (Advances in psychology 103), ed. by Jeanette Altarriba, 379–84. Amsterdam: North-Holland. DOI: 10.1016/S0166-4115(08)61673-8.

Avcu, Enes. 2018. Experimental investigation of the subregular hypothesis. *West Coast Conference on Formal Linguistics (WCCFL)* 35.77–86. Online: http://www.lingref.com/cpp/wccfl/35/paper3377.pdf.

Baer-Henney, Dinah, and Ruben van de Vijver. 2012. On the role of substance, locality, and amount of exposure in the acquisition of morphophonemic alternations. *Laboratory Phonology* 3.221–49. DOI: 10.1515/lp-2012-0013.

Bajec, Anton, et al. 2000. Slovar slovenskega knjižnega jezika. Online: http://bos.zrc-sazu.si/sskj.html.

Becker, Michael; Nihan Ketrez; and Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87.84–125. DOI: 10.1353/lan.2011.0016.

Becker, Michael, and Jonathan Levine. 2013. Experigen—an online experiment platform. Online: http://becker.phonologist.org/experigen.

Beguš, Gašper. 2017. Effects of ejective stops on preceding vowel duration. *The Journal of the Acoustical Society of America* 142.2168–84. DOI: 10.1121/1.5007728.

Beguš, Gašper. 2018. *Unnatural phonology: A synchrony-diachrony interface approach*. Cambridge, MA: Harvard University dissertation. Online: http://nrs.harvard.edu/urn-3:HUL.InstRepos:40050094.

Beguš, Gašper. 2019. Post-nasal devoicing and the blurring process. *Journal of Linguistics* 55.689–753. DOI: 10.1017/S002222671800049X.

Beguš, Gašper. 2020. Estimating historical probabilities of natural and unnatural processes. *Phonology* 37.515–49. DOI: 10.1017/S0952675720000263.

Bell, Alan. 1970. *A state-process approach to syllabicity and syllabic structure*. Stanford, CA: Stanford University dissertation.

Bell, Alan. 1971. Some patterns of the occurrence and formation of syllabic structure. *Working Papers on Language Universals* 6.23–138.

Blevins, Juliette. 2004. *Evolutionary phonology*. Cambridge: Cambridge University Press.

Broselow, Ellen. 2018. Laryngeal contrasts in second language phonology. *Phonological typology*, ed. by Larry M. Hyman and Frans Plank, 312–40. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110451931-009.

Bybee, Joan. 2001. *Phonology and language use*. (Cambridge studies in linguistics 94.) Cambridge: Cambridge University Press.

Carpenter, Angela C. 2010. A naturalness bias in learning stress. *Phonology* 27.345–92. DOI: 10.1017/S0952675710000199.

Cathcart, Chundra A. 2015. A probabilistic model of evolutionary phonology. *North East Linguistic Society (NELS)* 45.145–50.

Clark, Eve V., and Melissa Bowerman. 1986. On the acquisition of final voiced stops. *The Fergusonian impact: In honor of Charles A. Ferguson on the occasion of his 65th birthday. Vol. 1: From phonology to society*, ed. by Joshua A. Fishman, 51–68. Berlin: De Gruyter. DOI: 10.1515/9783110873641-006.

Coetzee, Andries W., and Rigardt Pretorius. 2010. Phonetically grounded phonology and sound change: The case of Tswana labial plosives. *Journal of Phonetics* 38.404–21. DOI: 10.1016/j.wocn.2010.03.004.

Cross, Ian. 2012. Cognitive science and the cultural nature of music. *Topics in Cognitive Science* 4.668–77. DOI: 10.1111/j.1756-8765.2012.01216.x.

Davidson, Lisa. 2016. Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics* 54.35–50. DOI: 10.1016/j.wocn.2015.09.003.

Davidson, Lisa. 2018. Phonation and laryngeal specification in American English voiceless obstruents. *Journal of the International Phonetic Association* 48.331–56. DOI: 10.1017/S0025100317000330.

de Lacy, Paul. 2006a. *Markedness: Reduction and preservation in phonology*. (Cambridge studies in linguistics.) Cambridge: Cambridge University Press.

de Lacy, Paul. 2006b. Transmissibility and the role of the phonological component: A theoretical synopsis of evolutionary phonology. *Theoretical Linguistics* 32.185–96. DOI: 10.1515/TL.2006.012.

de Lacy, Paul, and John Kingston. 2013. Synchronic explanation. *Natural Language & Linguistic Theory* 31.287–355. DOI: 10.1007/s11049-013-9191-y.

Dickens, Patrick. 1984. The history of so-called strengthening in Tswana. *Journal of African Languages and Linguistics* 6.97–125. DOI: 10.1515/jall.1984.6.2.97.

Do, Youngah, and Jonathan Havenhill. 2021. Production and substantive bias in phonological learning. *Proceedings of the 2020 Annual Meeting on Phonology*. DOI: 10.3765/amp.v9i0.4925.

Do, Youngah; Elizabeth Zsiga; and Jonathan Havenhill. 2016. Naturalness and frequency in implicit phonological learning. Paper presented at the 90th annual meeting of the Linguistic Society of America, Washington, DC, January 7–10.

Efron, Bradley. 1979. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 7.1–26. Online: http://www.jstor.org/stable/2958830.

Efron, Bradley. 1987. Better bootstrap confidence intervals. *Journal of the American Statistical Association* 82.171–85. DOI: 10.1080/01621459.1987.10478410.

Ettlinger, Marc; Kara Morgan-Short; Mandy Faretta-Stutenberg; and Patrick C. M. Wong. 2016. The relationship between artificial and second language learning. *Cognitive Science* 40.822–47. DOI: 10.1111/cogs.12257.

Ferdinand, Vanessa; Simon Kirby; and Kenny Smith. 2019. The cognitive roots of regularization in language. *Cognition* 184.53–68. DOI: 10.1016/j.cognition.2018.12.002.

Finley, Sara. 2012. Typological asymmetries in round vowel harmony: Support from artificial grammar learning. *Language and Cognitive Processes* 27.1550–62. DOI: 10.1080/01690965.2012.660168.

Finley, Sara, and William Badecker. 2009. Artificial language learning and feature-based generalization. *Journal of Memory and Language* 61.423–37. DOI: 10.1016/j.jml.2009.05.002.

Garrett, Andrew, and Keith Johnson. 2013. Phonetic bias in sound change. *Origins of sound change: Approaches to phonologization*, ed. by Alan C. L. Yu, 51–97. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199573745.003.0003.

Gauvain, Mary. 1995. Thinking in niches: Sociocultural influences on cognitive development. *Human Development* 38.25–45. DOI: 10.1159/000278297.

Glewwe, Eleanor. 2017. Substantive bias in phonotactic learning: Positional extension of an obstruent voicing contrast. Paper presented at the 53rd meeting of the Chicago Linguistic Society, Chicago, May 25–27.

Glewwe, Eleanor; Jesse Zymet; Jacob Adams; Rachel Jacobson; Anthony Yates; Ann Zeng; and Robert Daland. 2017. Substantive bias and word-final voiced obstruents: An artificial grammar learning study. Paper presented at the 92nd annual meeting of the Linguistic Society of America, Salt Lake City, January 4–7.

Greenberg, Joseph H. 1978. Diachrony, synchrony, and language universals. *Universals of human language, vol. 1: Method & theory*, ed. by Joseph H. Greenberg, 61–92. Stanford, CA: Stanford University Press.

Griffiths, Thomas L.; Michael L. Kalish; and Stephan Lewandowsky. 2008. Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363.3503–14. DOI: 10.1098/rstb.2008.0146.

Hansson, Gunnar Ólafur. 2008. Diachronic explanations of sound patterns. *Language and Linguistics Compass* 2.859–93. DOI: 10.1111/j.1749-818X.2008.00077.x.

Hayden, Rebecca E. 1950. The relative frequency of phonemes in General-American English. *Word* 6.217–23. DOI: 10.1080/00437956.1950.11659381.

Hayes, Bruce. 1999. Phonetically-driven phonology: The role of optimality theory and inductive grounding. *Functionalism and formalism in linguistics, vol. 1: General papers*, ed. by Michael Darnell, Edith Moravscik, Michael Noonan, Frederick J. Newmeyer, and Kathleen Wheatley, 243–85. Amsterdam: John Benjamins.

Hayes, Bruce, and Donca Steriade. 2004. Introduction: The phonetic bases of phonological markedness. *Phonetically based phonology*, ed. by Bruce Hayes, Robert Kirchner, and Donca Steriade, 1–33. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486401.001.

Hayes, Bruce, and Tanya Stivers. 2000. Postnasal voicing. Los Angeles: University of California, Los Angeles, ms. Online: http://linguistics.ucla.edu/people/hayes/Phonet/NCPhonet.pdf, accessed May 9, 2018.

Hayes, Bruce, and James White. 2013. Phonological naturalness and phonotactic learning. *Linguistic Inquiry* 44.45–75. DOI: 10.1162/LING_a_00119.

Hayes, Bruce; Kie Zuraw; Péter Siptár; and Zsuzsa Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85.822–63. DOI: 10.1353/lan.0.0169.

Haynie, Hannah J., and Claire Bowern. 2016. Phylogenetic approach to the evolution of color term systems. *Proceedings of the National Academy of Sciences* 113.13666–71. DOI: 10.1073/pnas.1613666113.

Hyman, Larry M. 1975. *Phonology: Theory and analysis*. New York: Holt, Rinehart & Winston.

Hyman, Larry M. 2001. The limits of phonetic determinism in phonology: *NC revisited. *The role of speech perception in phonology*, ed. by Elizabeth Hume and Keith Johnson, 141–86. San Diego: Academic Press.

Hyman, Larry M. 2013. Enlarging the scope of phonologization. *Origins of sound change: Approaches to phonologization*, ed. by Alan C. L. Yu, 3–28. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199573745.003.0001.

Iverson, Gregory K., and Joseph C. Salmons. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12.369–96. DOI: 10.1017/S0952675700002566.

Kessler, Brett, and Rebecca Treiman. 1997. Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language* 37.295–311. DOI: 10.1006/jmla.1997.2522.

Kiparsky, Paul. 1995. The phonological basis of sound change. *The handbook of phonological theory*, ed. by John Goldsmith, 640–70. Oxford: Blackwell.

Kiparsky, Paul. 2006. Amphichronic program vs. evolutionary phonology. *Theoretical Linguistics* 32.217–36. DOI: 10.1515/TL.2006.015.

Kiparsky, Paul. 2008. Universals constrain change; change results in typological generalizations. *Linguistic universals and language change*, ed. by Jeff Good, 23–53. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199298495.003.0002.

Kirby, Simon; Mike Dowman; and Thomas L. Griffiths. 2007. Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences* 104.5241–45. DOI: 10.1073/pnas.0608222104.

Kirby, Simon; Tom Griffiths; and Kenny Smith. 2014. Iterated learning and the evolution of language. *Current Opinion in Neurobiology* (Special issue: *Communication and language*, ed. by Michael S. Brainard and W. Tecumseh Fitch) 28.108–14. DOI: 10.1016/j.conb.2014.07.014.

Kluender, Keith R.; Randy L. Diehl; and Beverly A. Wright. 1988. Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics* 16.153–69. DOI: 10.1016/S0095-4470(19)30480-2.

Kong, Eun Jong; Mary E. Beckman; and Jan Edwards. 2012. Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of Phonetics* 40.725–44. DOI: 10.1016/j.wocn.2012.07.002.

Kümmel, Martin. 2007. *Konsonantenwandel*. Wiesbaden: Reichert.

Kuo, Li-Jen. 2009. The role of natural class features in the acquisition of phonotactic regularities. *Journal of Psycholinguistic Research* 38.129–50. DOI: 10.1007/s10936-008-9090-2.

Labov, William. 1994. *Principles of linguistic change, vol. 1: Internal factors*. Oxford: Blackwell.

Labov, William. 2001. *Principles of linguistic change, vol. 2: Social factors*. Oxford: Wiley-Blackwell.

Ladefoged, Peter, and Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Marvin, Tatjana; Jure Derganc; Samo Beguš; and Saba Battelino. 2018. Word selection in the Slovenian sentence matrix test for speech audiometry. *Zbornik konference jezikovne tehnologije in digitalna humanistika*, ed. by Darja Fišer and Andrej Pančur, 181–87. Ljubljana: Znanstvena založba Filozofske fakultete v Ljubljani. Online: http://www.sdjt.si/wp/dogodki/konference/jtdh-2018/zbornik-jtdh-2018/.

Mielke, Jeff. 2018. PBase: A database of phonological patterns. Online: http://pbase.phon.chass.ncsu.edu.

Mines, M. Ardussi; Barbara F. Hanson; and June E. Shoup. 1978. Frequency of occurrence of phonemes in conversational English. *Language and Speech* 21.221–41. DOI: 10.1177/002383097802100302.

Moran, Steven; Daniel McCloy; and Richard Wright (eds.) 2014. *PHOIBLE online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Online: https://phoible.org/.

Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25.83–127. DOI: 10.1017/S0952675708001413.

Moreton, Elliott, and Joe Pater. 2012a. Structure and substance in artificial-phonology learning, part I: Structure. *Language and Linguistics Compass* 6.686–701. DOI: 10.1002/lnc3.363.

Moreton, Elliott, and Joe Pater. 2012b. Structure and substance in artificial-phonology learning, part II: Substance. *Language and Linguistics Compass* 6.702–18. DOI: 10.1002/lnc3.366.

Moreton, Elliott; Joe Pater; and Katya Pertsova. 2017. Phonological concept learning. *Cognitive Science* 41.4–69. DOI: 10.1111/cogs.12319.

Moreton, Elliott, and Katya Pertsova. 2017. Implicit and explicit processes in phonotactic learning. *Proceedings of the Boston University Conference on Language Acquisition (BUCLD)* 40.277–90.

Morley, Rebecca L. 2012. The emergence of epenthesis: An incremental model of grammar change. *Language Dynamics and Change* 2.59–97.

Morley, Rebecca L. 2015. Can phonological universals be emergent? Modeling the space of sound change, lexical distribution, and hypothesis selection. *Language* 91.e40–e70. DOI: 10.1353/lan.2015.0019.

Myers, Scott. 2002. Gaps in factorial typology: The case of voicing in consonant clusters. Austin: University of Texas at Austin, ms. Online: http://roa.rutgers.edu/article/view/519.

Nisbett, Richard E., and Ara Norenzayan. 2002. Culture and cognition. *Stevens' handbook of experimental psychology*, 3rd edn., ed. by Hal Pashler, 561–97. Oxford: Wiley-Blackwell. DOI: 10.1002/0471214426.pas0213.

Ohala, John J. 1981. The listener as a source of sound change. *Chicago Linguistic Society (Parasession on language and behavior)* 17(2).178–203.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. *The production of speech*, ed. by Peter F. MacNeilage, 189–216. New York: Springer. DOI: 10.1007/978-1-4613-8202-7_9.

Ohala, John J. 1993. The phonetics of sound change. *Historical linguistics: Problems and perspectives*, ed. by Charles Jones, 237–78. London: Longman.

Ohala, John J. 1997. Aerodynamics of phonology. *Proceedings of the 4th Seoul International Conference on Linguistics (SICOL)*, 84–91.

Ohala, John J. 2006. Speech aerodynamics. *Encyclopedia of language & linguistics*, 2nd edn., ed. by Keith Brown, 684–89. Oxford: Elsevier.

Pater, Joe. 2004. Austronesian nasal substitution and other NC̦ effects. *Optimality theory in phonology: A reader*, ed. by John J. McCarthy, 271–89. Malden, MA: Blackwell.

Picard, Marc. 1994. *Principles and methods in historical phonology: From Proto-Algonkian to Arapaho*. Montreal: McGill-Queen's University Press.

Prince, Alan, and Paul Smolensky. 2004 [1993]. *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell. [Revision of technical report 2, Rutgers University Center for Cognitive Science, 1993.]

Pycha, Anne; Pawel Nowak; Eurie Shin; and Ryan Shosted. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. *West Coast Conference on Formal Linguistics (WCCFL)* 22.101–14.

Rafferty, Anna N.; Thomas L. Griffiths; and Marc Ettlinger. 2013. Greater learnability is not sufficient to produce cultural universals. *Cognition* 129.70–87. DOI: 10.1016/j.cognition.2013.05.003.

Reali, Florencia, and Thomas L. Griffiths. 2009. The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition* 111.317–28. DOI: 10.1016/j.cognition.2009.02.012.

Sankoff, Gillian, and Hélène Blondeau. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83.560–88. DOI: 10.1353/lan.2007.0106.

Schloss, Karen B., and Stephen E. Palmer. 2011. Aesthetic response to color combinations: Preference, harmony, and similarity. *Attention, Perception, & Psychophysics* 73.551–71. DOI: 10.3758/s13414-010-0027-0.

Seidl, Amanda; Eugene Buckley; and Alejandrina Cristià. 2007. Complexity trumps naturalness. Paper presented at the 81st annual meeting of the Linguistic Society of America, Anaheim, CA, January 4–7, 2007.

Simons, Gary F., and Charles D. Fennig (eds.) 2018. *Ethnologue: Languages of the world*. 21st edn. Dallas: SIL International. Online: https://www.ethnologue.com/, accessed October 13, 2018.

Singh, Sadanand, and John W. Black. 1966. Study of twenty-six intervocalic consonants as spoken and recognized by four language groups. *The Journal of the Acoustical Society of America* 39.372–87. DOI: 10.1121/1.1909899.

Skoruppa, Katrin, and Sharon Peperkamp. 2011. Adaptation to novel accents: Feature-based learning of context-sensitive phonological regularities. *Cognitive Science* 35.348–66. DOI: 10.1111/j.1551-6709.2010.01152.x.

Smith, Caroline L. 1997. The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics* 25.471–500. DOI: 10.1006/jpho.1997.0053.

Solé, Maria-Josep; Larry M. Hyman; and Kemmonye C. Monaka. 2010. More on post-nasal devoicing: The case of Shekgalagari. *Journal of Phonetics* 38.604–15. DOI: 10.1016/j.wocn.2010.09.002.

Steriade, Donca. 2001. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. Los Angeles: University of California, Los Angeles, ms.

Suhadolc, Barbara. 2013. Statistična analiza slovenskih besedil [Statistic analysis of Slovenian texts]. Ljubljana: University of Ljubljana master's thesis. Online: http://eprints.fri.uni-lj.si/2157/.

Šuštaršič, Rastislav; Smiljana Komar; and Bojan Petek. 1995. Slovene. *Journal of the International Phonetic Association* 25.86–90. DOI: 10.1017/S0025100300005211.

Tessier, Anne-Michelle. 2012. Testing for OO-faithfulness in the acquisition of consonant clusters. *Language Acquisition* 19.144–73. DOI: 10.1080/10489223.2012.660552.

Toporišič, Jože. 2004. *Slovenska slovnica*. Maribor: Obzorja.

van de Vijver, Ruben, and Dinah Baer-Henney. 2014. Developing biases. *Frontiers in Psychology* 5:634. DOI: 10.3389/fpsyg.2014.00634.

Verdonik, Darinka, and Ana Zwitter Vitez. 2011. Slovenski govorni korpus Gos. Ljubljana: Trojina, zavod za uporabno slovenistiko. Online: http://www.korpus-gos.net/.

Wang, William S.-Y., and John Crawford. 1960. Frequency studies of English consonants. *Language & Speech* 3.131–39. DOI: 10.1177/002383096000300302.

White, James. 2014. Evidence for a learning bias against saltatory phonological alternations. *Cognition* 130.96–115. DOI: 10.1016/j.cognition.2013.09.008.

White, James. 2017. Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a P-map bias. *Language* 93.1–36. DOI: 10.1353/lan.2017.0001.

White, James, and Megha Sundara. 2014. Biased generalization of newly learned phonological alternations by 12-month-old infants. *Cognition* 133.85–90. DOI: 10.1016/j.cognition.2014.05.020.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30.945–82. DOI: 10.1207/s15516709cog0000_89.

Yu, Alan C. L. 2011. On measuring phonetic precursor robustness: A response to Moreton. *Phonology* 28.491–518. DOI: 10.1017/S0952675711000236.

University of California, Berkeley
Department of Linguistics
1203 Dwinelle Hall #2650
Berkeley, CA 94720-2650
[begus@berkeley.edu]

[Received 24 June 2020;
revision invited 23 August 2020;
revision received 11 February 2021;
revision invited 10 April 2021;
revision received 4 May 2021;
accepted pending revisions 27 May 2021;
revision received 8 June 2021;
accepted 8 June 2021]