Zero-weighted constraints in Noisy Harmonic Grammar[1]

Bruce Hayes
UCLA

Aaron Kaplan
University of Utah

January 2023

To appear in *Linguistic Inquiry*

Keywords: Noisy Harmonic Grammar, harmonic bounding, clipping, zero weight

1. Introduction: Exploring the Behaviors of Constraint-based Theories

Constraint-based theories in linguistics have long served a dual purpose: they not only provide a framework for analysis of linguistic phenomena, but they also have their own formal properties that make broad yet precise predictions about what can occur in human languages. As these theories have evolved, a tradition of research has emerged aiming to extract and test such predictions. Such work began with classical Optimality Theory (OT; Prince and Smolensky 1993) and Harmonic Grammar (HG; Legendre, Miyata, and Smolensky 1990) and has continued with the probabilistic descendants of these frameworks: Stochastic OT (Boersma 1998), Maximum Entropy grammars (Goldwater and Johnson 2003), and Noisy Harmonic Grammar (Boersma and Pater 2016).[2]

We follow this research tradition here, examining some previously-unnoticed behaviors of Noisy Harmonic Grammar, hereafter NHG. All of these behaviors concern constraints that have been assigned a weight of zero. Our findings in brief are as follows. (a) As first noted by Flemming (2021), zero weighted constraints are not "turned off," but continue to influence the candidate evaluation. As we will show, these influences can be substantial; indeed, in contexts of harmonic bounding, a zero-weighted constraint can completely rule out candidates that violate it. (b) In the simplest version of NHG, zero-weighted constraints can reward, rather than penalize, candidates that violate them.

[1] We would like to thank Edward Flemming, the *LI* reviewers, and the participants in the UCLA Phonology Seminar (Fall 2021) for helpful input on this squib.

[2] For diagnostic work on classical OT, see Prince and Smolensky 1993, Prince 1997, Anttila et al. 2008, and Mai and Baković 2020; for classical HG, Bane and Riggle 2010, Jesney 2016, and Pater 2016; for Stochastic OT, Zuraw and Hayes 2017; for MaxEnt, Jesney 2007, Anttila and Magri 2018, and Flemming 2021; for Noisy Harmonic Grammar, Jesney 2007, Hayes 2017, Kaplan 2021, and Flemming 2021, 2022.

These predictions depend on particular choices that must be made concerning negative weights and harmonic bounding. At the end of the squib we discuss whether these are good predictions, how they might be tested empirically, and how rival frameworks differ.

2. Review of Noisy Harmonic Grammar

We first review the basics of NHG; for the original presentation see Boersma and Pater 2016.[3] As in OT, NHG assumes a set of inputs, a set of candidates for each input, a set of constraints, and a procedure for selecting winners based on the constraint violations. Unlike in OT, the constraints are not ranked but weighted; i.e. assigned real numbers reflecting their strength. Every candidate is assigned a Harmony score, computed by multiplying weights by violation counts for every constraint, then summing across all constraints. In the original, nonstochastic version of Harmonic Grammar, there is a unique winner, which is the candidate with the lowest Harmony score.[4]

Noisy Harmonic Grammar, in contrast, adds a probabilistic element to the evaluation of candidates; thus it is suitable for the analysis of the many linguistic phenomena that are gradient (see e.g. Bod, Hay and Jannedy 2003, Coetzee and Pater 2011). Representative work employing NHG includes Jesney and Tessier 2009, Coetzee and Kawahara 2013, and Kaplan 2021. To achieve probabilistic outputs, NHG adds a random noise value to each constraint's weight.[5] This value, ε, is drawn from a Gaussian distribution that has a mean of 0 and a standard deviation that we set here at 1. We will refer to the sum of the base weight and noise factor ε as the *perturbed* weight; a separate set of perturbed weights is employed at each evaluation time, or application of the grammar. To illustrate, tableau (1) has two candidates and two conflicting constraints weighted at 4.95 and 4.[6]

[3] Prepublication versions of this work date from 2008, though the essential reference to "clipping" discussed below is found only in later versions.

[4] The latter reflects our choice of sign conventions: we use positive weights for constraints that assess a penalty and positive integers to count violations; other authors have adopted different conventions.

[5] The text gives the classical version of the theory as proposed by Boersma and Pater; for alternatives see Hayes 2017 and Flemming 2017, 2021.

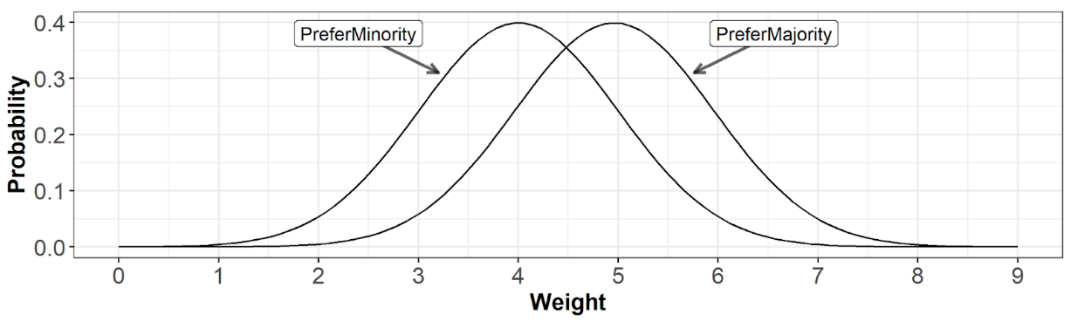
[6] The value 4.95 is rounded for convenience; the exact value for deriving .75/.25 probability is closer to 4.95387. The calculations for this squib, which were carried out with a combination of Excel and OTSoft 2.6 (Hayes, Tesar, and Zuraw 2021), may be inspected in the Supplementary Materials.

(1) *Two conflicting constraints with similar weights, deriving .75/.25 probability*

		PreferMajority ($w = 4.95 + \varepsilon_1$)	PreferMinority ($w = 4 + \varepsilon_2$)	Harmony	probability
Input1	Majority		1	$4 + \varepsilon_2$	.75
	Minority	1		$4.95 + \varepsilon_1$	.25

At each evaluation time, the constraint weights are perturbed by the values ε_1 and ε_2. Across evaluation times, perturbed weights fall into two overlapping Gaussian distributions, centered at the base weights 4.95 and 4, as shown in figure 1:

Figure 1. Probability distributions for the perturbed weights in Tableau (1)



In this case, since each candidate has just one violation of one constraint, the Harmony value of each candidate is equal to the perturbed weight of the constraint that it violates. Because of these distributions, the Minority candidate will have the greater Harmony penalty more often than not, but on some trials the opposite outcome will obtain. With a 0.95 difference in base weights, the Majority candidate wins 75% of the time. More generally, the difference in probability depends on the difference in base weights.

3. Harmonic Bounding and Negative Weights

It is a widely held, though not universal, view that it is desirable for a constraint-based framework to impose the property of *harmonic bounding* (Prince and Smolensky 1993:156): any candidate that has a strict superset of the violations of any other candidate can never win.[7] Work advocating the use of theories that maintain harmonic bounding includes Anttila and Magri 2018, Mai and Baković 2020, and Kaplan 2021; skeptical work that actually relies on the absence of harmonic bounding for linguistic analysis includes Hayes and Wilson 2008, Kaplan 2011, and Hayes and Schuh 2019.

[7] We address here only simple harmonic bounding, where a losing candidate is bounded by a single rival; for collective harmonic bounding see e.g. Samek-Lodovici and Prince 1999.

Classical OT is a clear example of a theory that respects harmonic bounding. Nonstochastic HG also respects harmonic bounding, provided that negative weights are prohibited. Such weights turn constraint violations into rewards (Keller 2000:314), reversing harmonic bounding relationships (Pater 2009). This is illustrated in (2) and (3) below, in which the candidate Bounder always defeats Bounded, which has a superset of its violations. Since the constraint INDIFFERENT penalizes the two candidates equally, any positive weight assigned to PREFERBOUNDER will ensure the defeat of Bounded.

(2) *Nonstochastic HG: harmonic bounding holds when weights are positive*

	INDIFFERENT (*w* = 6)	PREFERBOUNDER (*w* = 0.5)	Harmony
☞ Bounder	1		6
Bounded	1	1	6.5

If, in contrast, PREFERBOUNDER's weight is negative, as in (3), the violation becomes a credit for the candidate Bounded, and this candidate will win, contrary to our original intent. (For the case of zero-weighted PREFERBOUNDER, which creates a tie, see section 4.1.2 below.)

(3) *Nonstochastic HG: harmonic bounding fails if weights can be negative*

	INDIFFERENT (*w* = 6)	PREFERBOUNDER (*w* = −0.5)	Harmony
Bounder	1		6
☞ Bounded	1	1	5.5

Turning to NHG, we find that it is not enough simply to require the base weights to be positive, since even when this is so, the *perturbed* weight may be negative. This, too, defeats harmonic bounding: in the case of (2), noise will often cause PREFERBOUNDER to bear a negative perturbed weight, yielding a positive probability for Bounded. This scenario is shown in tableau (4), which includes the noise factors and the computed probabilities.

(4) *Failure of harmonic bounding in one version of Noisy Harmonic Grammar*

	INDIFFERENT ($w = 6 + \varepsilon_1$)	PREFERBOUNDER ($w = 0.5 + \varepsilon_2$)	Harmony	Probability
Bounder	1		$6 + \varepsilon_1$	0.691
Bounded	1	1	$6 + \varepsilon_1 + 0.5 + \varepsilon_2$	0.309

In short, if harmonic bounding is to be maintained, the theory is in need of some repair to preclude this scenario.

To this end, Boersma and Pater (2016), following Keller 2000, suggest that weights that have been perturbed into the negative zone should undergo *clipping*, receiving by fiat the value zero instead of the assigned negative value; see also Magri 2015, addressing learnability for clipping. We agree that clipping can help enforce harmonic bounding, but it is insufficient on its own; observe that in (4), clipping a negative perturbed weight for PREFERBOUNDER to zero would create a tie between the two candidates, each receiving the Harmony score $6 + \varepsilon_1$. Hence something needs to be said about how to deal with ties in NHG.

Boersma and Pater's suggestion is that when a tie occurs, the winner should be picked at random from among the tied candidates. This is in one respect a sensible choice, since it helps generate incorrect winners that can guide the learning of constraint weights; see Jesney and Tessier 2011. However, it does not solve the harmonic bounding problem. In (4), PREFERBOUNDER will go below zero 0.309 of the time when outputs are derived. Assuming clipping, these cases will result in a tie. Then, assuming the random-selection method of tie resolution, the candidate Bounded will be picked $0.309/2 \approx 0.154$ of the time. The harmonic bounding problem remains, for we have only halved the probability of the Bounded candidate rather than reduced it to zero.

We propose instead that ties should result in *trial cancellation*: when an evaluation yields tied winners, no output is chosen, and we move on to the next trial. Thus all cases in which a harmonically bounded candidate might win are excluded: in (4) either PREFERBOUNDER is perturbed to a positive value, so Bounder wins in the ordinary way, or PREFERBOUNDER is perturbed to a negative value, in which case clipping creates a tie, which invokes trial cancellation.[8] In the latter case, the evaluation must be attempted anew until trial cancellation is not triggered. In the case of (4), there will be plenty of

[8] If PreferBounder is perturbed exactly to zero, trial cancellation is again invoked, and harmonic bounding is preserved.

cases over time in which PREFERBOUNDER is perturbed to a positive weight, yielding a verdict. The set of uncancelled trials will respect harmonic bounding.[9]

In sum, we suggest that NHG can be prevented from generating harmonically bounded winners by deploying a combination of clipping and trial cancellation.[10] We will explore the consequences of adopting (or not adopting) these procedures, as we turn to our main topic, the effects of zero-weighted constraints.

4. Assessing the Effects of Zero-Weighted Constraints

Consider now an augmented version of tableau (1): we add an additional Input2 as well as a zero-weighted constraint to be called MAJORITYHELPER. MAJORITYHELPER is violated by the Minority candidate in Input2 only. Intuitively, MAJORITYHELPER is an "ally" of PREFERMAJORITY, because with a sufficiently positive weight it would indeed help the Majority candidate. However, our focus is on the special case where MAJORITYHELPER bears a weight of zero. The scenario is given in (5).

(5) *Assessing the effect of a zero-weighted constraint aligned with PREFERMAJORITY*

		PREFERMAJORITY ($w = 4.95 + \varepsilon_1$)	PREFERMINORITY ($w = 4 + \varepsilon_2$)	MAJORITYHELPER ($w = 0 + \varepsilon_3$)	Harmony	*P*
Input1	Majority		1		$4 + \varepsilon_2$	.75
	Minority	1			$4.95 + \varepsilon_1$	.25
Input2	Majority		1		$4 + \varepsilon_2$	?
	Minority	1		1	$4.95 + \varepsilon_1 + \varepsilon_3$	?

As in (1), the weights of PREFERMAJORITY and PREFERMINORITY will result in 75% probability for the Majority candidate for Input1. These weights are sufficiently high that the effect of clipping is negligible for these constraints. We now calculate the probabilities for Input2; as it turns out, the outcome depends on whether we employ

[9] One other conceivable case is if the two candidates have exactly the same violations, which would lead to an infinite loop as trial cancellation is repeatedly invoked. We assume that in such cases the candidates should be assigned equal probability by fiat. Note that these cases do not involve harmonic bounding.

[10] The approach described here could be described as "sample-then-clip." An alternative pointed out by an LI reviewer is "clip-then-sample": the perturbed weights are sampled from a normal distribution truncated at zero. Under this approach, Trial Cancellation is unnecessary. The effects described in section 4.1 get stronger, because the perturbed weights clipped to zero under sample-then-clip never arise.

clipping or not (hence the question marks in (5)). The two cases must be considered separately.

4.1 Clipping Imposed: Zero Weights, but Non-Zero Effects

Under clipping we have determined that the probability of the Majority candidates in (5) are 0.75 for Input1 and 0.812 for Input2. Thus, even though MAJORITYHELPER has a zero weight, it has a modest empirical effect (per Flemming 2021); the zero weight does not turn it off.

Why should this be so? The reason becomes apparent if one examines the perturbed weights assigned to MAJORITYHELPER across a series of trials. Because we are assuming clipping, on about half of the trials its weight will be zero and the outcome will be determined by the other constraints. On the other half of the trials, MAJORITYHELPER will have a positive weight, which will boost the Harmony penalty against the Minority candidate, thus raising the Majority candidate's probability. In other words, clipping creates a weight distribution for MAJORITYHELPER that consists exclusively of non-negative values. In this sense, MAJORITYHELPER was not really zero-weighted in the first place.

4.1.1 Effect Size

For purposes of actual linguistic analysis, it is useful to know how large the effect just described can be; if these effects are so small as to be empirically indetectible, there is little point in changing the theory to avoid them. In fact, we find that in (5) if MAJORITYHELPER is violated just once, the largest effect is about 0.106, occurring when PREFERMAJORITY is outweighed by PREFERMINORITY by 0.058. If MAJORITYHELPER can be violated more than once, the effect can become very large, approaching 0.5; this is found where PREFERMINORITY is weighted far above PREFERMAJORITY and MAJORITYHELPER is violated a great number of times, placing *P*(Majority) near zero for Input1 and near 0.5 for Input2. These effects are illustrated in sections 4 and 5 of the Supplementary Materials.

4.1.2 More on effect size: the case of harmonic bounding

Moving beyond the scenario of (5), we note a different case where the effect of a zero-weighted constraint is as large as 0.5, namely under conditions of harmonic bounding. In (6) we give a revised version of (4) in which the weight of PREFERBOUNDER is set at zero. With clipping and trial cancellation in effect, the probability of the Bounded candidate comes out as zero. Had PREFERBOUNDER not been in the grammar, then the candidates Bounding and Bounded would have received 50/50 probability, per fn. 9.

(6) *Harmonic bounding resulting from a zero-weighted constraint*

	INDIFFERENT ($w = 6 + \varepsilon_1$)	PREFERBOUNDER ($w = 0 + \varepsilon_2$)	Harmony	*P*
☞ Bounding	1		$6 + \varepsilon_1$	1
Bounded	1	1	$6 + \varepsilon_1 + \varepsilon_2$	0

The effect of PREFERBOUNDER arises as follows. Whenever ε_2 is negative, it triggers clipping, so that the perturbed weight of PREFERBOUNDER is zero. This creates a tie, and the trial is canceled. When ε_2 is zero, again there is a tie and the trial is canceled. When ε_2 is positive, the Bounding candidate wins; hence, it wins in all non-canceled trials. The example is the clearest possible illustration of the point that giving a constraint a zero weight is not the same as removing it from the grammar.

4.2 Clipping Not Imposed: Violators are Rewarded

We return to the schematic example (5), which was intended to test the effects of zero-weighted MAJORITYHELPER. We just showed that when clipping is in effect, MAJORITYHELPER helps the Majority candidate. In contrast, when clipping is turned off, it emerges that MAJORITYHELPER actually ends up *hurting* the Majority candidate for Input2; this candidate receives a probability of 0.709, somewhat lower than the 0.750 obtained for Input1. In other words, absent clipping, NHG predicts that the effect of a constraint can be the opposite of what the analyst may have intended in formulating it.

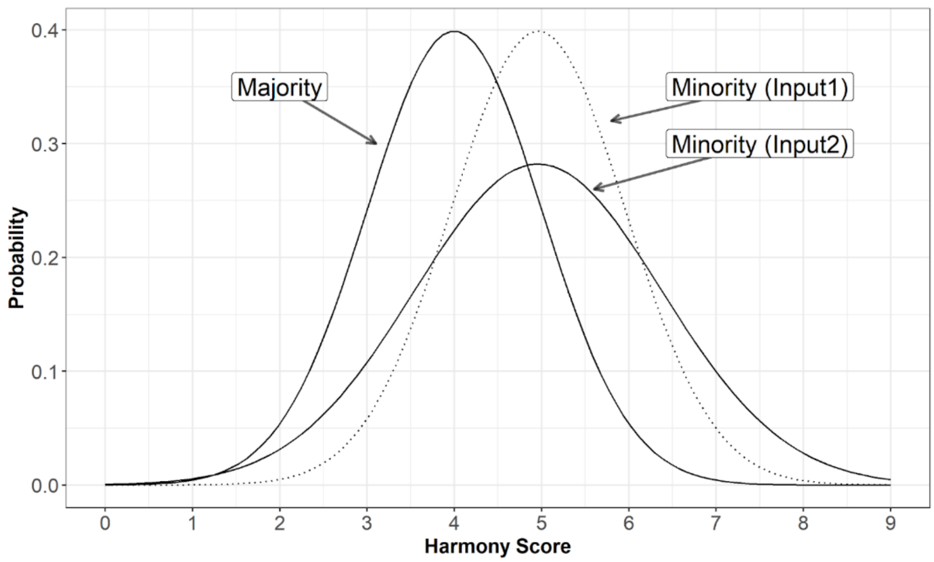
To understand how this reversal can arise, we return to the basic mechanisms of NHG. In figure 1 we showed how the two overlapping probability distributions for the perturbed weights of PREFERMINORITY and PREFERMAJORITY result in a 0.75 probability for the Majority candidate of Input1. For Input2, the effect of MAJORITYHELPER on the probability of the Minority candidate must also be included. To do this, we calculate the probability distribution of its Harmony penalty, which can be read off of tableau (5) as follows:

(7) *Probability distribution of Harmony value for tableau (5), Input2, Minority candidate*

$$\begin{aligned} H(\text{Minority}) &= [w(\text{PREFERMAJORITY}) + \varepsilon_1] + [w(\text{MAJORITYHELPER}) + \varepsilon_3] \\ &= [\,4.95 + \varepsilon_1] + [\,0 + \varepsilon_3] \end{aligned}$$

This probability distribution is the sum of two Gaussians, one of them arising from PREFERMAJORITY, the other from MAJORITYHELPER. The sum of two Gaussians *x* and *y* is also a Gaussian, whose mean is the sum of the means (here, 4.95 + 0), and whose standard deviation is obtained from the formula $\sqrt{\sigma_x{}^2 + \sigma_y{}^2}$; in this case $\sqrt{1^2 + 1^2} \cong$ 1.414. This Harmony distribution can be compared with the one for the Minority candidate for Input1, which has the same mean (4.95), but a standard deviation of just 1. Thus, MAJORITYHELPER is in effect a source of "pure noise," which *broadens* the probability distribution of the Harmony penalty assigned to the Minority candidate. This is shown in figure 2, where the original "unbroadened" distribution is shown with a dotted line.

Figure 2. MAJORITYHELPER broadens the probability distribution of Harmony



It is this broadening that increases the number of cases in which the perturbed Harmony of the Minority candidate is less than that of the Majority candidate, leading to the counterintuitive shift. In particular, the left tail of the combined distribution reaches deeper into the territory of the Majority candidate's distribution, resulting in more reversed outcomes, so that across trials the Minority candidate becomes more probable. If MAJORITYHELPER is violated multiple times, the Harmony distribution of the Minority candidate becomes even broader, and the observed effect is increased.[11]

As before, we check what is the largest possible effect. When MAJORITYHELPER is violated once, the largest reversed effect occurs when PREFERMAJORITY is weighted 1.56 higher than PREFERMINORITY; this changes Majority's probability from 0.865 to 0.816, i.e. by 0.049. When MAJORITYHELPER is violated a large number of times, the maximum probability reduction approaches 0.5, as shown in the Supplementary Materials.

Lastly, we note that the effects observed in this section are found even when the weight of MAJORITYHELPER is not actually zero, but merely small. To give an example, we have calculated that for (5), if there is no clipping, MAJORITYHELPER has a reversed

[11] For more on the broadening of Harmony distributions and their empirical consequences, see Flemming 2021, 2022.

effect with weights ranging from zero to 0.214; above this value it helps the Majority candidate (see Supplementary Materials).

5. Discussion

5.1 Empirical Predictions

To review, the discussion above establishes that zero-weighted constraints in NHG can be non-inert, and this non-inertness takes two forms. If clipping is assumed, a zero-weighted constraint can reduce the probability of candidates that violate it, to a degree that varies according to other factors. If clipping is not assumed, a zero-weighted constraint can increase the probability of candidates that violate it, again with variation depending on other factors. We consider next what language data might bear on the predictions made above. We offer two cases.

(a) *Reversal*. NHG without clipping predicts that there should be constraints that assess a penalty when they are strong but provide a reward when they are weak. For Markedness, this means that the very same configuration can be evaluated anywhere from slightly good to very bad. A Faithfulness constraint can encourage faithfulness when strong, or discourage Faithfulness when weak. We know of no cases of either kind, and are therefore skeptical of NHG without clipping.

(b) *No turn-off*. Even with clipping in place, NHG still predicts that zero-weighted constraints cannot be turned off; they alter output probabilities by non-trivial amounts. This prediction interacts in an intriguing way with the hypothesis of the Universal Constraint Set (Prince and Smolensky 1993:2), for it implies that certain frequency distributions should never occur. For instance, suppose that the constraint *ROLO (no nonhigh rounded vowels; Kaun 1995, 2004) is in the universal constraint set. Then (barring the introduction of ad hoc constraints) one cannot describe a system in which optional rounding harmony applies with equal frequency to low vowels (OA → OO) and high vowels (UI → UU). Giving *ROLO a weight of zero will not accomplish this.

5.2 Alternatives

The potential predictions just outlined distinguish NHG from alternative approaches to constraint-based stochastic grammar. While this squib is not the place for extended theory-comparison,[12] we can address how these alternatives compare with NHG regarding the issues considered here. The points of comparison are (a) whether the theory

[12] Some of the literature contributing to the debates is the following. MaxEnt, for: Hayes and Wilson 2008, Zuraw and Hayes 2017, Hayes and Schuh 2019, Smith and Pater 2020, Flemming 2021, Hayes 2022; against: Anttila and Magri 2018, Anttila, Borgeson, and Magri 2019, Kaplan 2021. Stochastic OT, for: Boersma and Hayes 2001; against: Keller and Asudeh 2002, McPherson and Hayes 2016, Zuraw and Hayes 2017; Classical NHG, for: Zuraw and Hayes 2017, Kaplan 2021; against: Flemming 2021, Hayes 2022; Exponential NHG: underexplored, no literature after Boersma and Pater 2016 of which we are aware.

assigns zero probability to harmonically-bounded candidates, (b) whether zero-weighted constraints (or the closest analog) are turned off, (c) whether there are “reversal” effects as defined above, and (d) whether a zero-weighted constraint can act as a harmonic bounder. We cover the behaviors of three theories; for reasons of space we cannot present full demonstrations of these points here, though they are straightforward.

MaxEnt grammars use a mathematical formula (Goldwater and Johnson 2003, ex. (1)) that translates Harmony into probability. In this theory: (a) harmonically bounded candidates can receive positive probability (though never the highest probability); (b) Giving a constraint zero weight turns it off completely, hence (c) the reversal syndrome described above cannot occur; and (d) zero-weighted constraints are likewise totally ineffective even in a harmonic bounding configuration.

Exponential NHG (Boersma and Pater 2016) adds a further step to the Harmony computation: base weights are perturbed, and the result is then exponentiated, which always creates a positive value, even if the perturbed weight was negative. Therefore, the candidate competition never references negative weights. Properties: (a) harmonic bounding is respected. (b) No constraint is ever turned off, though the influence of a constraint on the outcome (cf. (5)) can approach zero as its weight tends toward $-\infty$. (c) Since exponentiated weights are never negative, the reversal syndrome cannot arise. (d) Even constraints with highly negative base weights can create harmonic bounding.

Stochastic OT (Boersma 1998) has a different structure from the other theories: constraint-specific “ranking values” are perturbed and sorted to create classical OT rankings, each employed for just one evaluation time. With respect to properties (a)-(d), the theory behaves qualitatively just like Exponential NHG.

In sum, there are multiple frameworks currently under study that differ in the predictions about properties (a)-(d). In principle future empirical research can use these properties to help distinguish among these frameworks, as well as frameworks yet to be devised.

References

Anttila, Arto, Vivienne Fong, Štefan Beňuš, and Jennifer Nycz. 2008. Variation and opacity in Singapore English consonant clusters. *Phonology* 25.2:181–216. doi: https://doi.org/10.1017/S0952675708001462

Arto Anttila, Scott Borgeson, and Giorgio Magri. 2019. Equiprobable mappings in weighted constraint grammars. In *SIGMORPHON 2019: Proceedings of the 16th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, Florence. doi: https://doi.org/10.48550/arXiv.1907.05839

Anttila, Arto, and Giorgio Magri. 2018. Does MaxEnt overgenerate? Implicational universals in Maximum Entropy grammar. In *Proceedings of the Annual Meetings on Phonology*, vol. 5. doi: https://doi.org/10.3765/amp.v5i0.4260

Bane, Max, and Jason Riggle. 2010. The typological consequences of weighted constraints. In *Proceedings of the 45th Annual Meeting of the Chicago Linguistic Society*.

Boersma, Paul. 1998. Functional phonology: Formalizing the interactions between articulatory and perceptual drives. Doctoral dissertation, University of Amsterdam.

Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32:45–86. doi: https://doi.org/10.1162/002438901554586

Boersma, Paul, and Joe Pater. 2016. Convergence properties of a gradual learning algorithm for Harmonic Grammar. In McCarthy and Pater 2016, 389–434.

Bod, Rens, Jennifer Hay, and Stefanie Jannedy, eds. 2003. *Probabilistic Linguistics*. Cambridge: MIT Press.

Coetzee, Andries W., and Shigeto Kawahara. 2013. Frequency biases in phonological variation. *Natural Language and Linguistic Theory* 31:47–89. doi: https://doi.org/10.1007/s11049-012-9179-z

Coetzee, Andries and Joe Pater. 2011. The place of variation in phonological theory. In *The Handbook of Phonological Theory* (2nd ed.), ed. by John Goldsmith, Jason Riggle, and Alan Yu, 401–431. Malden, MA: Wiley-Blackwell.

Flemming, Edward. 2017. Stochastic Harmonic Grammars as random utility models. Poster presented at the Annual Meeting in Phonology, New York University.

Flemming, Edward. 2021. Comparing MaxEnt and Noisy Harmonic Grammar. *Glossa: A journal of general linguistics* 6:1–42. doi: https://doi.org/10.16995/glossa.5775

Flemming, Edward. 2022. MaxEnt vs. Noisy Harmonic Grammar. Paper given at the 8th International Conference on Phonology and Morphology, Seoul.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a Maximum Entropy model. In *Proceedings of the Stockholm Workshop on Variation within Optimality Theory,* ed. by Jennifer Spenader, Anders Eriksson, and Östen Dahl, 111–120. Stockholm: Stockholm University.

Hayes, Bruce. 2017. Varieties of Noisy Harmonic Grammar. In *Proceedings of AMP 2016*, ed. by Karen Jesney, Charlie O'Hara, Caitlin Smith, and Rachel Walker. doi: https://doi.org/10.3765/amp.v4i0.3997

Hayes, Bruce. 2022. Deriving the Wug-shaped curve: A criterion for assessing formal theories of linguistic variation. *Annual Review of Linguistics* 8:474-494. doi: https://doi.org/10.1146/annurev-linguistics-031220-013128

Hayes, Bruce, Bruce Tesar, and Kie Zuraw. 2021. OTSoft 2.6. Software, online at linguistics.ucla.edu/people/hayes/otsoft/.

Hayes, Bruce, and Russell Schuh. 2019. Metrical structure and sung rhythm of the Hausa rajaz. *Language* 95:e253–e299. doi: https://doi.org/10.1353/lan.2019.0043

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379–440. doi: https://doi.org/10.1162/ling.2008.39.3.379

Jesney, Karen. 2007. The locus of variation in weighted constraint grammars. Paper given at the *Workshop on Variation, Gradience and Frequency in Phonology,* Stanford, CA.

Jesney, Karen. 2016. Positional constraints in Optimality Theory and Harmonic Grammar. In McCarthy and Pater 2016, 176–220.
Jesney, Karen, and Anne-Michelle Tessier. 2009. Gradual learning and faithfulness: Consequences of Ranked vs. Weighted Constraints. In Muhammad Abdurrahman, Anisa Schardl & Martin Walkow, eds., *Proceedings of the 38th Meeting of the North East Linguistic Society* (*NELS* 38). Amherst, MA: GLSA.
Jesney, Karen, and Anne-Michelle Tessier. 2011. Biases in Harmonic Grammar: The road to restrictive learning. *Natural Language and Linguistic Theory* 29.1:251–290. doi: https://doi.org/10.1007/s11049-010-9104-2
Kaplan, Aaron. 2011. Variation through Markedness Suppression. *Phonology* 28.3:331–370. doi: https://doi.org/10.1017/S0952675711000200
Kaplan, Aaron. 2021. Categorical and gradient ungrammaticality in optional processes. *Language* 97:703–731. doi: https://doi.org/10.1353/lan.2021.0062
Kaun, Abigail. 1995. The Typology of Rounding Harmony: An Optimality Theoretic Approach. Doctoral dissertation, UCLA.
Kaun, Abigail. 2004. The typology of rounding harmony. In *Phonetically-Based Phonology*, ed. by Bruce Hayes, Robert Kirchner, and Donca Steriade, 87–116. Cambridge: Cambridge University Press.
Keller, Frank. 2000. Gradience in grammar: Experimental and computational aspects of degrees of grammaticality. Doctoral dissertation, The University of Edinburgh.
Keller, Frank, and Ash Asudeh. 2002. Probabilistic learning algorithms and Optimality Theory. *Linguistic Inquiry* 33:225–244. doi: https://doi.org/10.1162/002438902317406704
Legendre, Géraldine, Yoshiro Miyata, and Paul Smolensky. 1990. Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, 388–395. Cambridge, MA: Lawrence Erlbaum.
Mai, Anna, and Eric Baković. 2020. Cumulative constraint interaction and the equalizer of OT and HG. *Proceedings of the 2019 Annual Meeting on Phonology*. doi: https://doi.org/10.3765/amp.v8i0.4678
Magri, Giorgio. 2015. How to keep the HG weights non-negative: The truncated Perceptron reweighing rule. *Journal of Language Modelling* 3.2:345–375. doi: https://doi.org/10.15398/jlm.v3i2.115
McCarthy, John, and Joe Pater, eds. 2016. *Harmonic Grammar and Harmonic Serialism*. London: Equinox Press.
McPherson, Laura, and Bruce Hayes. 2016. Relating application frequency to morphological structure: The case of Tommo So vowel harmony. *Phonology* 33:125–167. doi: https://doi.org/10.1017/S0952675716000051
Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33:999–1035. doi: https://doi.org/10.1111/j.1551-6709.2009.01047.x
Pater, Joe. 2016. Universal Grammar with weighted constraints. In McCarthy and Pater 2016, 1–46.

Prince, Alan, and Paul Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. Ms., Rutgers University, New Brunswick and University of Colorado, Boulder. Published 2004, Malden, MA: Blackwell.

Prince, Alan. 1997. Stringency and anti-Paninian hierarchies. Lecture given at LSA Institute, Cornell University.

Samek-Lodovici, Vieri, and Alan Prince. 1999. Optima. Rutgers Optimality Archive ROA-363.

Smith, Brian W., and Joe Pater. 2020. French schwa and gradient cumulativity. *Glossa: a journal of general linguistics* 5:1–33. doi: https://doi.org/10.5334/gjgl.583

Zuraw, Kie, and Bruce Hayes. 2017. Intersecting constraint families: An argument for Harmonic Grammar. *Language* 93:497–548. doi: https://doi.org/10.1353/lan.2017.0035

Bruce Hayes
University of California, Los Angeles
bhayes@humnet.ucla.edu

Aaron Kaplan
University of Utah
a.kaplan@utah.edu