Phonology **34** (2017). Supplementary materials

On the replicator dynamics of lexical stress: accounting for stress-pattern diversity in terms of evolutionary game theory

Andreas Baumann
Nikolaus Ritt
University of Vienna

Supplementary materials

Appendix A: Replicator dynamics

Replicator dynamics (Hofbauer & Sigmund 1998: 67f, Nowak 2006: 45f) model frequency-dependent selection (Heino *et al.* 1998) as a dynamical system in continuous time. The rates of change of the frequencies x and y, i.e. their increase or decrease per time unit, are given in (18).

(18) a. $\dot{x} = x(f_1(x,y) - \varphi(x,y))$
b. $\dot{y} = y(f_2(x,y) - \varphi(x,y))$

Here, $f_1(x,\mathrm{y}) = a_{11}x + a_{12}y$ and $f_2(x,y) = a_{21}x + a_{22}y$ denote the fitness of S_1 players and S_2 players respectively, and $\varphi(x,y) = xf_1(x,y) + yf_1(x,y)$ denotes the average fitness for a given composition of frequencies (x,y). The growth rates of both player types are determined by (a) the frequency of individuals adopting a strategy and (b) the difference between strategy-specific fitness and average fitness.

Since $y = 1 - x$, it is sufficient to restrict the model to the frequency of S_1 players. For a 2×2 payoff matrix $\mathrm{A} = (a_{ij})$ the two-dimensional system above can be reduced to (19).

(19) $\dot{x} = x(1-x)((a_{11} + a_{22} - a_{12} - a_{21})x + a_{12} - a_{22})$

Dynamical equilibria, i.e. situations in which the distribution of strategies among players does not change in one or the other direction, are of particular interest. Mathematically, this is the case if $\dot{x} = 0$. Then the population is in an equilibrium, denoted by $\hat{x}$. For the equation above, there are three potential equilibria: two 'pure' ones $\hat{x}_1 = 1$ (all players play strategy S_1, e.g. all words take initial stress) and $\hat{x}_2 = 0$ (all players play strategy S_2, e.g. all words take final stress), as well as an 'internal equilibrium', shown in (20), as long as $\hat{x}_{int}$ is positive.

(20) $\hat{x}_{int} = \frac{a_{22} - a_{12}}{a_{11} + a_{22} - a_{12} - a_{21}}$

In this configuration, some words take initial stress and others final stress. If population is at equilibrium, its composition will not change unless it is perturbed by some system-external factor. Whether or not the population returns to the equilibrium depends on its 'stability'. In replicator dynamics, stable equilibria are referred to as 'evolutionarily stable strategies'.

Assessing the stability of $\hat{x}_1$, $\hat{x}_2$ and $\hat{x}_{int}$ is simple: if $a_{11} > a_{21}$ and $a_{12} > a_{22}$, then $\hat{x}_1$ is stable, $\hat{x}_2$ is unstable and there is no internal equilibrium $\hat{x}_{int}$. That is, if initially stressed words (S_1) do better than finally stressed ones (S_2) both when they combine with other initially stressed words (S_1) and when they combine with finally stressed ones (S_2), then the lexicon will eventually contain only initially stressed words. In other words, the dynamics result in a population of S_1 players only. The opposite is true if $a_{11} < a_{21}$ and $a_{12} < a_{22}$. Then $\hat{x}_1$ is unstable while $\hat{x}_2$ is stable, and this leads to a population consisting exclusively of S_2 players (i.e. final stress throughout the lexicon).

If $a_{11} > a_{21}$ and $a_{12} < a_{22}$, then both $\hat{x}_1$ and $\hat{x}_2$ are stable, and an internal equilibrium (representing a specific mix of initially stressed words and finally stressed ones) does exist. However, it is unstable. This means that if initially stressed words do better than finally stressed ones when they combine with initially stressed words, but worse when they combine with finally stressed ones, the lexicon will eventually also come to contain only one of the two types.

The only system in which an internal equilibrium $\hat{x}_{int}$ will be stable is one in which $a_{11} < a_{21}$ and $a_{12} > a_{22}$. In that case, it is always better for both players to have opposing strategies, i.e. initially stressed words do better than finally stressed ones when combining with finally stressed words, but worse when combing with initially stressed ones. Under such conditions, the dynamics will inevitably result in a stable mix of S_1 and S_2 players, i.e. the lexicon will turn out as a historically stable mix of initially stressed words and finally stressed ones. The proportion of words with initial stress (technically: the fraction of S_1 players) in a stable equilibrium depends on the quotient above. Note that the long-term dynamics of an evolutionary game are fully determined by the entries in the pay-off matrix.

Appendix B: Discrete probability distributions: the 2-simplex

Probability distributions $\pi = (\pi_1, \ldots, \pi_n)$ among $n \in \mathbb{N}^+$ disjoint categories $C_1, \ldots, C_n$ can be visualised as single points in a so-called simplex, or more precisely, a $(n-1)$-simplex, where all points π fulfil $0 \leq \pi_1, \ldots, \pi_n \leq 1$ and $\pi_1 + \ldots + \pi_n = 1$, since all entries of π are probabilities distributed among n categories. In the case of three categories, i.e. $n = 3$, the set of all possible probability distributions thus defines a two-dimensional surface, the 2-simplex, embedded in three-dimensional space (Fig. 5). The closer one moves towards, say, the corner defined by $\pi_1 = 1$, or equivalently $\pi = (1, 0, 0)$, the more dominant category C_1 becomes.

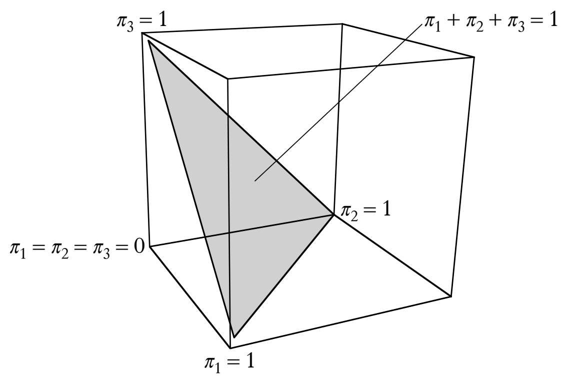

Figure 5

The 2-simplex (grey) as a two-dimensional surface in three-dimensional space. The coordinates of all points $\pi = (\pi_1, \pi_2, \pi_3)$ in the 2-simplex are non-negative and sum to 1.

Appendix C: Derivation of the pay-off matrix

Since the pay-offs for all entries in the matrix $A^{\pi,\varphi,\alpha}$ are determined in the same way, we demonstrate the procedure for only one of them. For that purpose we chose $a_{11}^{\pi,\varphi,\alpha}$, which represents the pay-offs for an encounter of two initially stressed words $['\sigma\sigma]_A$ and $['\sigma\sigma]_B$.

In context C_1, the sequences $['\sigma\sigma]_A[\sigma]['\sigma\sigma]_B$ and $['\sigma\sigma]_B[\sigma]['\sigma\sigma]_A$ are formed. Both produce a single violation (a lapse) and incur a rhythmicity score of $\varphi(1) = \varphi_{max} - \Delta\varphi_{max}$. In the first sequence, $['\sigma\sigma]_A$ receives a pay-off of $\alpha(\varphi_{max} - \Delta\varphi_{max})$, while $['\sigma\sigma]_B$ gets $(1-\alpha)(\varphi_{max} - \Delta\varphi_{max})$. In the second sequence, $['\sigma\sigma]_A$ gets $(1-\alpha)(\varphi_{max} - \Delta\varphi_{max})$, while $['\sigma\sigma]_B$ gets $\alpha(\varphi_{max} - \Delta\varphi_{max})$. In total, $['\sigma\sigma]_A$ receives $\alpha(\varphi_{max} - \Delta\varphi_{max}) + (1-\alpha)(\varphi_{max} - \Delta\varphi_{max}) = \varphi_{max} - \Delta\varphi_{max}$. The same holds for $['\sigma\sigma]_B$.

In context C_2, the sequences $['\sigma\sigma]_A['\sigma]['\sigma\sigma]_B$ and $['\sigma\sigma]_B['\sigma]['\sigma\sigma]_A$ are formed. Again, both produce a single violation (in this case a clash) and incur a rhythmicity score of $\varphi(1) = \varphi_{max} - \Delta\varphi_{max}$. As above, both $['\sigma\sigma]_A$ and $['\sigma\sigma]_B$ get $\varphi_{max} - \Delta\varphi_{max}$.

In context C_3, the sequences $['\sigma\sigma]_A['\sigma\sigma]_B$ and $['\sigma\sigma]_B['\sigma\sigma]_A$ are formed. Neither produces a violation, so that both incur scores of $\varphi(0) = \varphi_{max}$. Hence both $['\sigma\sigma]_A$ and $['\sigma\sigma]_B$ get $\alpha\varphi_{max} + (1-\alpha)\varphi_{max} = \varphi_{max}$.

Since the proportion of encounters in each of the three contexts reflects distribution π, the entry of the pay-off matrix can be expressed as in (21).

(21) $a_{11}^{\pi,\varphi,\alpha} = \pi_1(\varphi_{max} - \Delta\varphi_{max}) + \pi_2(\varphi_{max} - \Delta\varphi_{max}) + \pi_3\varphi_{max}$

ADDITIONAL REFERENCE

Heino, Mikko, Johan A. J. Metz & Veijo Kaitala (1998). The enigma of frequency-dependent selection. *Trends in Ecology and Evolution* **13**. 367–370.