Semantics & Pragmatics Volume 10, Article 17: 1–23, 2017
https://doi.org/10.3765/sp.10.17

Analyzing imperfective games*

Igor Yanovich
DFG Center for Advanced Studies
"Words, Bones, Genes and Tools"

Submitted 2016-09-26 / Decision 2016-10-24 / Revisions received 2017-08-23 / Decision 2017-09-30 / Accepted 2017-10-16 / Early access 2017-12-20 / Final publication 2020-04-06

Abstract Deo 2015 is the first study applying mathematically explicit evolutionary analysis to a specific semantic-change phenomenon, namely the progressive-imperfective diachronic cycle. However, Deo's actual results do not match completely the empirical observations about that cycle. Linguistic communities passing through the cycle often employ, in the synchrony, a single common type of progressive-imperfective grammar. In Deo's modeling results, however, two of the grammars never get shared by nearly all the population, including the grammar with the obligatory use of progressive marking in semantically progressive contexts, as in Present-Day English. This paper improves on that wrong prediction. The crucial modeling decision enabling the improvement is switching from the assumption of infinite speaker population to the more realistic, but harder to analyze finite population setting. The finite-population version of Deo's model derives stages where at many time points, all or almost all speakers share the same grammar. Interestingly, two different *a priori* reasonable types of trajectories with that feature emerge, depending on the parameter settings. These two trajectory types constitute novel empirical predictions regarding the shape of the cycle generated by (the proposed extension of) Deo's model.

Keywords: evolutionary modeling, semantic change, progressive, imperfective, genetic drift

* I am very grateful to Roland Mühlenbernd and especially to Ashwini Deo, as well as to the anonymous reviewers for S&P and editors Louise McNally and Michael Franke. Their comments made the paper substantially better. Financial support by the Alexander von Humboldt Foundation and by DFG under Project FOR 2237: "Words, Bones, Genes, Tools: Tracking Linguistic, Cultural, and Biological Trajectories of the Human Past", is hereby acknowledged with gratitude.

The progressive-imperfective cycle of semantic change is as follows. It starts with grammar (a) where only one linguistic form *X* is used for both progressive and imperfective meanings. Present-Day Bulgarian is an example of stage (a).[1] In the next stage (b), an optional progressive marker *Y* is innovated. Early Modern English is an example of this: in Shakespeare's early 17-century *Tragedy of Hamlet*, Polonius says to Hamlet "What do you *read*?" clearly meaning to ask what Hamlet *was reading* at the moment. Thus the morphological imperfective could express the semantic progressive. But the *be -ing* form was already available for signaling semantic progressivity explicitly. Indeed, in the same play Ophelia uses the innovative form: "as I *was sewing* in my closet, Lord Hamlet ... comes before me". At stage (c), the younger marker *Y* becomes obligatory for progressive meanings, while older *X* is still retained for imperfective ones. Present-Day English is an example of this stage: there are no longer speakers who would produce Polonius's phrase with a progressive meaning. Finally, at stage (d) *X* dies out altogether, and *Y* expresses both the imperfective and the progressive. Modern Turkish is apparently currently moving into that stage. (d) is just like the original stage (a), only with a new undifferentiated progressive-imperfective form. The cycle may start anew when a still newer optional progressive form is innovated.

The current paper is a follow-up improving on Deo 2015, an evolutionary-modeling study on the progressive-imperfective cycle of semantic change we just sketched. Deo's original model derives the shifting through the stages of the cycle (a)→(b)→(c)→(d), but does not predict that there will be many times when all or almost all speakers of a language would share the same progressive-imperfective grammar. This is contrary to what we observe. For example, in Present-Day English, we hardly have any native speakers who could use the simple imperfective *reads* with the meaning of the progressive *is reading* (grammar (b)). Nor do we have speakers who never ever use the simple imperfective form (grammar (d)). Present-Day English thus features virtually 100% of speakers carrying grammar (c), and hardly any speakers with grammars (b) and (d). But in Deo's original results, the peak frequency of (c) is 68%, with grammars (a) and (b) each at 1%, and grammar (d) at 30% in the speaker population at that moment.

The purpose of this contribution is to show that modeling the cycle is much more successful if we work with finite-population models, as opposed

1 Throughout the paper, I will sloppily use designations like (a) to refer both to a stage in the cycle and to an individual speaker's grammar dominant at that stage.

to Deo's original infinite-population one. Section 1 discusses the usefulness and uses of evolutionary modeling for semantic change. Section 2 describes the original model of Deo 2015. In Section 3, I show that switching to a finite-population model helps to solve the problem. I describe two types of *a priori* reasonable change trajectories generated by Deo's substantive theory of the cycle in the finite-population setting, and argue that they constitute novel, sharper predictions on the shape of the actual cycle, and thus call for more fine-grained empirical data to either support or refute the model.

For those who would like to engage in evolutionary modeling exercises themselves, or to simply replicate the types of analyses reported here, this article has a longer companion piece available at the Semantics Archive at http://semanticsarchive.net/Archive/jAxYjUzY/. The ca. 80-page-long companion paper discusses modeling in a more step-wise manner, and explains some of the simple techniques that allow one to better understand the behavior of an evolutionary model.

Modeling was implemented in R, a free software package for statistical computing with convenient visualization capabilities, R Core Team 2016. Annotated source code is available at the Semantics Archive together with the companion piece.

1 Why evolutionary modeling for semantic change?

A familiar approach to semantic change frames it in terms of broad stages. For each historical stage, we identify the forms present, their ranges of meanings, and the change pathways that connect the form-meaning pairings from different stages. As an example, we may so describe the evolution of English *be going to* construction (see Eckardt 2006: Ch. 4, a.o.): at the Late Middle English stage, *be going to* can only express directional meanings, while at the Late Modern English stage, it is frequently used as a futurate. This is broadly analogous to how a paleobiologist could describe a particular sequence of ecosystems in terms of the species inhabiting them at different times and their evolutionary relations: one species, the directional *be going to*, gave rise to another one, futurate *be going to*, which became much more frequent, and currently overshadows the still existing directional construction. But how does a population of speakers actually move between such stages? The synchronic characterization of the broad stages does not by itself explain the dynamic part of the process. Evolutionary modeling for semantic change, as applied in Deo 2015 to the progressive-imperfective cycle, attempts to provide an explicit dynamic component, complementing the essentially synchronic understanding of the individual stages.

The need to connect the synchrony with the diachrony is not unique to linguistics. To use an example from biology, lactase persistence is the ability of humans to digest lactose in milk beyond early childhood. Only about a third of the world population are lactase-persistent, and the geographical distribution of such individuals is highly skewed. The spread of lactase persistence is likely a recent event associated with the spread of milking of domestic animals. The genetic picture is complex, with several different mutations creating the lactase persistence effect (see Jones et al. 2013, Swallow 2003, Burger et al. 2007, a.o.) There are many dynamic questions that we may want to ask that go far beyond describing the current distribution of lactase persistence, and its likely absence in the early human past. How early did lactase-persistent genetic variants arise? Where did it happen? What caused their spread? To answer them, we need a theory that studies how genetic distributions change. The discipline of *population genetics* aims at building such a theory and analyzes patterns of change in the distributions of genetic variants. In other words, population genetics is an application of evolutionary modeling to genetics.

Whether in biology or in linguistics, evolutionary modeling may occur at different levels of analysis:

i. **Sanity-check level**: can a given model of change generate the observed changes?

ii. **Range-of-predictions level**: what is the range of possible change trajectories that can be generated under different parameter values of our model?

iii. **Parameter-inference level**: can we discover the likely values of our model's parameters by fitting the model to fine-grained empirical data?

Deo (2015) analyzes the progressive-imperfective semantic-change cycle at the sanity-check level. She demonstrates that her substantive semantic theory of the progressive and the imperfective, together with specific assumptions about misacquisition of progressive-imperfective grammars and specific assumptions she makes about the evolutionary laws governing grammar spread and loss, together induce cyclic shifting behavior in a modeled population of speakers.

In this reply, I aim at both the sanity-check level, improving on Deo 2015 therein, and at the range-of-predictions level. I show below that there are two distinct types of change trajectories for the progressive-imperfective grammars that arise from Deo's general framework. As there is currently no sufficiently fine-grained empirical data to attempt parameter inference, that level remains beyond the scope of the present contribution.

The levels above are not, of course, unique to linguistics. Let's use lactase persistence as an example. At the sanity check level, we can ask whether the current distribution of lactase-persistence genomes could *in principle* arise under the assumption that being able to drink milk beyond childhood provides survival and reproduction advantages. At the level of the range of predictions, we can ask *how robustly* the predictions of sane models match the modern distribution of lactase-persistent genomes. Finally, at the parameter-inference level, we can attempt to *quantify the exact strength* of survival and reproduction advantages that lactase persistence brings.

A frequent worry is that evolutionary models for language change may be oversimplified and therefore useless. It is true that proposed models are often quite simple. But that same charge would apply to the study of biological evolution as well. The difficulty is that studying the mathematics of evolution is not easy. Population genetics textbooks (Gillespie 2004 is a concise and accessible example) start from a very simple model called the Wright-Fisher model, where generations do not overlap, the population is not stratified either geographically or socially, and the number of offspring for each individual is binomially distributed (this last assumption would hold for a species producing an extremely large number of potential offspring from which only a few survive into the next generation; think of fish laying eggs). Everyone knows that those assumptions are highly non-realistic for most species. Adopting more realistic assumptions may substantially change the model's behavior (see, for example, Der, Epstein & Plotkin 2011.) But despite that, the oversimplified model is often applied in practice. Why should then anybody apply it? First, simpler models may be analytically tractable where more complex ones would not be. We can use the simpler models to build our mathematical intuitions that we can subsequently apply to more complex models. Second, even very simple models can sometimes get the predictions almost exactly right. That would in turn mean that they capture the essence of the evolutionary process, and that the additional factors not modeled only play a limited role. The point here, of course, is not that simpler models are necessarily better: they can easily fail to be adequate. But they should not be dismissed out of hand simply for clearly being simpler than the reality.

2 Deo's evolutionary setup

Diachronic cycles of change are a common phenomenon arising in various parts of grammar. One of such cycles in semantic change concerns progressive and imperfective markers, described at the beginning of this note. Deo (2015) applies evolutionary modeling to that cycle, predicting the trajectories of change in terms of the shares of speakers employing one of the possible grammar types of the cycle. Deo's model consists of several components:

- Analysis of communication between speakers with (different or same) progressive-imperfective grammars, resulting in quantitative assessment of relative communicative efficiency;
- Assumptions about misacquisition of progressive-imperfective grammars by children;
- Rules that govern the dynamic behavior of a system defined by the above two components.

Deo provides an analysis of communication between speakers of progressive-imperfective grammars game-theoretically, through a particular variety of *signaling game*. We restrict our attention to the reductions of full grammars to their progressive-imperfective components. An implicit assumption here is that the rest of the grammar does not directly affect what happens within this subsystem. We assume two linguistic forms X and Y, and two types of meanings to convey: we write **phen**(omenal) for the narrow-progressive meaning, and **struc**(tural) for the narrow-imperfective meaning.[2] Speakers use X and Y to attempt to convey meanings **phen** and **struc** to each other. We assume that the context of each utterance, visible to both speaker and hearer, can be of two types: in context C_{phen}, it is more likely that the speaker would want to convey **phen**, and in C_{struc}, she would be more likely to convey **struc**.

Within this formalism, a progressive-imperfective grammar amounts to a game-playing strategy that determines how speakers choose their message X

2 The intuition behind the terms is that some meanings are about specific events, or *phenomena*, while others are about the *structure* of the world.

or Y dependent on the intended meaning and the context, and how hearers reconstruct the meaning **phen** or **struc** based on the heard form. Deo restricts attention to four such strategies $\langle S_i, H_i \rangle$, where $i \in \{(a), (b), (c), (d)\}$,[3] corresponding to the four stages of the cycle. They are given in Table 1.[4]

	Speaker				Hearer			
a		**struc**	$\Rightarrow$	X	C_{struc}		$\Rightarrow$	**struc**
		phen	$\Rightarrow$	X	C_{phen}		$\Rightarrow$	**phen**
b	C_{struc}	**struc**	$\Rightarrow$	X	C_{struc}	X	$\Rightarrow$	**struc**
		phen	$\Rightarrow$	Y		Y	$\Rightarrow$	**phen**
	C_{phen}		$\Rightarrow$	X	C_{phen}		$\Rightarrow$	**phen**
c		**struc**	$\Rightarrow$	X		X	$\Rightarrow$	**struc**
		phen	$\Rightarrow$	Y		Y	$\Rightarrow$	**phen**
d		**struc**	$\Rightarrow$	Y	C_{struc}		$\Rightarrow$	**struc**
		phen	$\Rightarrow$	Y	C_{phen}		$\Rightarrow$	**phen**

Table 1 Four grammars (that is, speaker-hearer strategy pairs) of Deo 2015

We are now ready to assemble what we may call Deo's Basic Imperfective Game **BImp** = $\langle F, M, C, G, p, ident, cost \rangle$, where messages $F = \langle X, Y \rangle$, meanings $M = \langle \textbf{struc}, \textbf{phen} \rangle$, contexts $C = \langle C_{struc}, C_{phen} \rangle$, grammars G are the four pairs in Table 1, p is a probability distribution over $M \times C$, and $ident$ is a function from $M \times M$ to $\{0, 1\}$ that models success in communication: if the hearer recovered the intended meaning from M, $ident$ returns 1, and it returns 0 otherwise. Finally, $cost$ is a function punishing particular strategies. Deo uses it to model pressure against richer grammars, in the way discussed below.

The game is played as follows. Nature selects a pair of meaning and context (that is, a member of $M \times C$) according to probability distribution

3 Deo's original labels for the same grammars/stages are cd, pcd, em, cd'.

4 For the stage and grammar (b), it is not completely clear if the correspondence holds empirically. Grammar (b) employs the novel optional progressive marker Y only when the context favors the narrow-imperfective meaning. I am not aware of evidence that would show that empirical speakers of the (b) stage indeed use their innovative progressive forms exclusively in contexts favoring narrow-imperfective readings, though such use wouldn't be surprising.

p. The speaker observes both elements, while the hearer only observes the context. The speaker sends a signal from F according to her grammar, which is one of those in G. The hearer uses her grammar, which may be the same or different from the speaker's, to reconstruct the meaning based on the signal and the context. If meaning recovery is successful, both speaker and hearer are assigned 1 by $ident$. If there is a mismatch, both are assigned 0. Finally, function $cost$ assigns a penalty from $[0, 1]$ to the speaker based on the complexity of her grammar. The sum of what the players get assigned by $ident$ and $cost$ forms their *payoffs* — a numerical measure of success in communication.

To make this general model more concrete, Deo makes the following assumptions. $cost$ is used to punish the speakers who employ an extra form by penalty k. So one-form grammars (a) and (d) are favored by $cost$ over more expressive two-form grammars (b) and (c). Regarding p, Deo assumes that $p(\langle \mathbf{phen}, C_{phen} \rangle) = p(\langle \mathbf{struc}, C_{struc} \rangle) = 0.45$, while $p(\langle \mathbf{phen}, C_{struc} \rangle) = p(\langle \mathbf{struc}, C_{phen} \rangle) = 0.05$. In words, both contexts are equiprobable; the probability of a disfavored meaning appearing in a context is 0.1.

Given this setup, we can straightforwardly calculate the *expected payoff* for each pair of grammars playing against each other.[5] These are given in Table 2. If a pair of players play against each other an infinite number of times, their average payoffs will be equal to those in the table. We can then reinterpret the table A in Table 2 as a part of the definition for Normal-Form Imperfective Game **NFImp** $= \langle G, A, k \rangle$. **NFImp** conveniently hides the fine structure of **BImp**, representing the information about the expected payoffs in a more compact form. **NFImp** is thus the result of Deo's substantive analysis of communication between speakers with progressive-imperfective grammars — the first component of her evolutionary model.

The ultimate job of this component in the evolutionary analysis is to tell us the expected payoff (=a measure of communicative efficiency) of a speaker with a given grammar in a population with certain proportions of speakers of different grammars. For example, suppose the current population is 80% categorial progressive users, with grammar (c), and 20% optional progressive users, with grammar (b). Then a (c) speaker in this population gets the expected payoff of $80\% * (1 - k/2) + 20\% * (0.75 - k/2)$, which is just the payoffs from the third row of A above weighted by the share of the respective

5 This is just the familiar probability-theoretic notion of expectation applied to payoffs. Namely, expected payoff is a sum of would-be payoffs in each condition weighted by the probability of the condition.

Grammars	a	b	c	d
a	0.9	0.9	0.7	0.9
b	$0.9 - \frac{k}{2}$	$0.95 - \frac{k}{2}$	$0.75 - \frac{k}{2}$	$0.7 - \frac{k}{2}$
c	$0.7 - \frac{k}{2}$	$0.75 - \frac{k}{2}$	$1 - \frac{k}{2}$	$0.7 - \frac{k}{2}$
d	0.9	0.7	0.7	0.9

Note that A is equal to the symmetric matrix representing the communicative success minus the matrix that punishes the two two-form grammars b and c with $k/2$.

Table 2 Deo's payoff matrix A, or the Normal-Form Imperfective Game

grammars in the population. This amounts to $0.95 - k/2$, while a similar calculation for (b) results in $0.79 - k/2$. Clearly, (c) speakers get considerably better payoffs in this population than (b) speakers, so it is better to be a (c) speaker in it. Generally, the more frequent strategy in the population will always get better and better payoffs as its share increases towards 100%: as a brief inspection of A will show, communicating with a speaker with the same grammar is always the best one could hope for in this model. This is, of course, a welcome result: intuitively, it should be easier to communicate with those who share your language. In Deo's evolutionary model, A determines *fitness*, in evolutionary terms.[6]

The second component of Deo's setup models misacquisition of grammars. Deo assumes that a child whose target grammar is i will sometimes misacquire it, with fixed probability. Misacquisition is then modeled as *mutation* changing the child's grammar type at a certain rate.

Building a grammar different from the one that was used to produce the observed input is a necessary prerequisite for grammatical change, whether it happens in language acquisition or in adulthood. But there are different ways to model it. One particularly important choice is whether misacquisition

6 How sensitive Deo's model is to the specific numbers in A? After all, the exact numbers will change slightly if we make different arbitrary choices about the frequency of two types of context, and the frequency of two types of meanings conditional on the context. Fortunately, the model is only moderately sensitive to the exact numbers. The features of A that are crucial for the evolutionary process's shape are: (i) each grammar is most efficient for communicating with speakers with the same grammar; (ii) two-form grammars (b) and (c) are better for communication than one-form grammars (a) and (d), as long as cost k is kept relatively low; (iii) between the two-form grammars, the obligatory-progressive (c) is more efficient since it is the only one that allows for perfect recovery of the speaker's meaning, with no guessing necessary. As one can see, these features are natural and intuitive, and do not depend on specific frequencies that Deo picks.

should depend on the current population state as a whole, or only on the target grammar of the potentially misacquiring individual.[7] Deo chooses the second option.[8] Some formal studies into language change concentrate on the misacquisition process itself (see especially Niyogi 2006.) In contrast, Deo 2015 abstracts away from the mechanics of misacquisition: she uses her analyst's judgement to hypothesize likely constant misacquisition rates organized into matrix Q, in Table 3. While those rates are not derived from empirical data, they can be falsified by future empirical work. Here, we follow Deo (2015) in accepting them as a working hypothesis.[9] It is easy to read off Q that Deo assumes that 6% of target-(a) speakers would switch to the innovative grammar (b) with an optional progressive, and so forth.[10] The

7 In a uniform parental population, these coincide, but we are also interested in cases where the population as a whole is not uniform — for example, because some change is already ongoing.

8 If one wishes to model misacquisition as dependent on the population state as a whole, it would no longer be appropriate to model it as mutation. One should then add a further misacquisition module to the fitness component; it will then consist of some weighted multiplication of the fitnesses/payoffs determined from **NPImp** given in Table 2, and additional misacquisition fitnesses.

9 A reviewer asks whether the parameters in Q could be inferred empirically, namely from a corpus. Unfortunately, this is only possible in the context of joint inference of all parameters of the model. Inferring misacquisition rates from empirical data would involve identifying an individual's target grammar. Under the interpretation employed by Deo where a generation is a biological generation, that would involve identifying a child's target grammar presumably from among the grammars of the surrounding adults. Under an alternative interpretation, cf. fn. 13, where "generations" refer to time-slices of the same individuals, the target grammar is simply the actual grammar of the same person in the preceding time-slice. So as a first step, we can record how frequently the actually used grammar differs from the likely target grammar. However, the rate of such difference crucially *cannot* be equated with the rates in Q! That is because the model involves more communicatively efficient grammars being adopted as targets more frequently than their current share in the population would suggest. So the actual amount of change in the shares of two grammars is affected by (i) the rates in Q, (ii) the communicative efficiency in A, and (iii) the size, social structure, and linguistic composition of the surrounding speaker population, as well as by the precise shape of the evolutionary laws. There is simply no observable quantity produced by the hypothesized process that depends only on the rates in Q, hence we cannot estimate Q from the data without estimating the whole model.

10 The only "independent" non-zero rates in Q are four rates $(a) \to (b)$, $(b) \to (a)$, $(b) \to (c)$, and $(c) \to (d)$. (If a rate is set to zero, this means Deo judged the corresponding misacquisition type impossible.) The four diagonal rates are obtained by subtracting all other rates in the row from 1. It is the relative values of the "independent" rates that are important. In particular, they determine the position of the equilibria of the system: the stable states of the population. If we scale all four independent rates proportionally, by multiplying by the

exact numbers in Q have only moderate significance: the important role of the matrix as a whole is to provide a push for the $(a) \to (b) \to (c) \to (d)$ progression of the cycle, and many combinations of rates will do it just fine, though for simplicity we stick with Deo's original matrix in this note.[11]

Strategies	*a*	*b*	*c*	*d*
a	0.94	0.06	0	0
b	0.02	0.91	0.07	0
c	0	0	0.97	0.03
d	0	0	0	1

Table 3 Deo's misacquisition (=mutation) matrix Q

By themselves, matrices A and Q above do not determine how a population of progressive-imperfective grammar speakers would develop over time. We need evolutionary laws that determine that with the help of A and Q — the third component of a full evolutionary model. Evolutionary laws are not primitives; they arise deterministically from particular substantial assumptions about the evolutionary process.[12]

Here are the assumptions that Deo introduces, explicitly or implicitly, which result in a particular evolutionary law she adopts, called (discrete-time) replicator-mutator dynamics. Each individual in the model represents a speaker, who is assumed to fix their progressive-imperfective segment of grammar after language acquisition. Generations do not overlap, and time is discrete: the individuals at time $(t+1)$ acquire their progressive-imperfective grammars from the individuals of time t.

Communicative efficiency affects how many "linguistic children" a given adult will transfer their grammar to: that number is proportional to the speaker's communicative efficiency relative to the other speakers' efficiency.

same constant, we mostly affect the speed at which the system progresses. (If we did not have the communicative efficiency component as in matrix A, the positions of the equilibria would not shift under moderate scaling. With A, the equilibria will shift because the relative strength of A vs. Q will have changed.)

11 For illustration purposes, SM 1 shows infinite-population trajectories under a number of combinations of the four independent rates. It is easy to see that (i) for the system to progress to (d), misacquisition rates $(b) \to (c)$ and $(c) \to (d)$ need to be relatively high, and (ii) under no conditions do intermediate grammars (b) and (c) reach anywhere close to 100% frequency.

12 Some authors present evolutionary laws as if they were primitives. But this only makes the underlying assumptions implicit.

It is thus $f_i(\bar{x})/\phi(\bar{x})$, where $\bar{x} = \langle x_a, x_b, x_c, x_d \rangle$ is the vector representing the current share of speakers of each grammar in the population, $f_i(\bar{x})$ is the expected payoff of speakers with grammar i in population $\bar{x}$, and $\phi(\bar{x})$ is the average expected payoff in the population. In the example above with $x_b = 20\%$ and $x_c = 80\%$, we have $\phi(\bar{x}) = 20\% * f_b(\bar{x}) + 80 * \% f_c(\bar{x})$. Using the payoffs from A in Table 2, we compute $f_b(\bar{x}) = 20\% * (0.95 - k/2) + 80\% * (0.75 - k/2) = 0.79 - k/2$ and $f_c(\bar{x}) = 20\% * (0.75 - k/2) + 80\% * (1 - k/2) = 0.95 - k/2$. From that, we get $\phi(\bar{x}) = 20\% * (0.79 - k/2) + 80\% * (0.95 - k/2) = 0.918 - k/2$. For concreteness, let's fix k at 0.01, as Deo does. Then $f_b(\bar{x})/\phi(\bar{x}) \approx 0.860$, while $f_c(\bar{x})/\phi(\bar{x}) \approx 1.035$. Thus each (b) speaker will serve as a target for roughly 0.86 children on average, while each (c) speaker would do so for roughly 1.035 new-generation individuals. The way the formulas work, it is guaranteed that the overall size of the new generation is equal to the old size, so it always remains constant.

After the target grammar i of a child is determined, mutation can still force the child to acquire a different grammar, according to the probabilities in matrix Q. We can thus think of grammar transmission as a two-step process: in the first step, the target grammar for each child is picked; in the second, it gets determined whether the child mutates or acquires the target grammar faithfully.

Next, Deo makes the assumption that the population is infinite. With the infinite population, the above considerations lead to a *deterministic* evolutionary law, called replicator-mutator (RM) dynamics. Finally, Deo assumes that both rates in Q and communicative efficiencies in A should be adopted 1-to-1 to represent *mutation* and *fitness* in the RM dynamics. This is not a necessary assumption, so when checking our predictions, it is useful to see what happens when the relative magnitude of forces recorded in A and Q is preserved, but their absolute effect is scaled by two constants α and β. This gives us the following version of the RM dynamics:[13]

13 The RM dynamics actually arises from many different sets of substantive assumptions, not just from the set used in Deo 2015. For more details on how to arrive from such sets to particular evolutionary laws, see Yanovich 2016: Sec. 6 or, for a more formal and general exposition, Sandholm 2010: Ch. 4. In particular, there is a readily available different interpretation of Deo's evolutionary law that allows change in the adulthood: under that alternative interpretation, individuals of the model are *time slices* of actual speakers, with the previous time slice as the "parent" of the current time slice. As those individuals proceed through discrete time, they can sometimes adopt the grammars of their more successfully communicating neighbors, or be shifted into speaking a different grammar by mutation. Instead of being tied to a biological generation, the generation of the model will then be of

(1) **Replicator-mutator discrete-time dynamics, fitness & mutation scaled**

$$x_i' = \sum_j x_j(1 - \alpha(1 - \tfrac{f_j(\bar{x})}{\phi(\bar{x})}))Q_{ji}',$$
$$\text{where } Q_{ji}' = \beta Q_{ji} \text{ for } j \neq i,\ \ Q_{jj}' = 1 - \textstyle\sum_{i \neq j} \beta Q_{ji}$$

With $\alpha = \beta = 1$, (1) reduces to the more familiar RM equations in (2). Note, however, that (1) is also the standard RM dynamics, only with A and Q rescaled.

(2) $$x_i' = \sum_j x_j \tfrac{f_j(\bar{x})}{\phi(\bar{x})} Q_{ji}, \quad \text{where } f_j(\bar{x}) = \textstyle\sum_k x_k A_{jk}, \text{ and } \phi(\bar{x}) = \textstyle\sum_k x_k f_k(\bar{x})$$

Why assume that the population is infinite? This makes the evolutionary laws become deterministic, which is convenient: it means we can compute *exactly*, non-probabilistically, the state of the population given any initial conditions and a set number of generations. The evolutionary process as defined above is a random process on the level of one individual. So in our previous example, any single (c) speaker will have a random integer number of "linguistic children", with number 1.035 simply being the *expected* number. If we average the number of children among a finite number of (c) parents, that might amount to, say, around 1.02 or around 1.05. At any rate, it will be a random number. But when the population is infinite, all randomness averages out, and 1.035 becomes also the *actual average* number of offspring per (c) speaker. Evolutionary trajectories thus become deterministic, and much easier to analyze than the tremendous space of probabilistically possible trajectories for finite populations.

No doubt that many of the assumptions of Deo's evolutionary model are not realistic. As we already discussed in the previous section, this by itself is not a good reason to abandon the model. Sometimes simple models based on unrealistic assumptions perform surprisingly well — in which case the simplifications built into them turn out to be very useful idealizations that help us to get a better grasp on the reality around us. It is the predictions of the model that need to be examined before we can decide whether the simplifications it makes are useful or counterproductive.

Deo's model correctly derives the progression of the cycle from (a) to (d). This is not insignificant: Yanovich 2016: Sec. 6.5 shows that keeping the same A and Q, but adopting a different set of *a priori* reasonable evolutionary laws miserably fails to derive any cyclic behavior.

relatively arbitrary length, while the fact that such generations are not overlapping will not be unrealistic anymore.

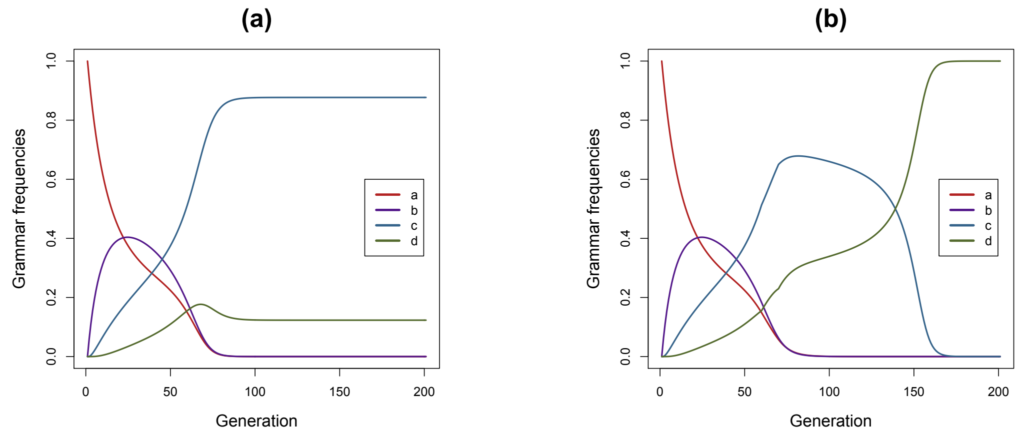

Figure 1 The evolutionary trajectories predicted by Deo 2015. (a) is a replica of Deo 2015: Fig. 3; (b), of Deo 2015: Fig. 4

However, there is a serious problem with the predictions that Deo's infinite-population model makes: it does not predict that there are times when (almost) all speakers share grammars (b) or (c). Figure 1(a) shows the predicted trajectories with Deo's A and Q under the standard RM dynamics. Figure 1(b) shows the trajectories under Deo's modification of the RM dynamics that gives an additional boost to grammar (d).[14] Neither trajectory set predicts a language like Present-Day English, where all speakers use the progressive obligatorily to express progressive meanings (grammar (c)), or like Early Modern English, where for all speakers, the *be -ing* form was available, but was not obligatory to express the semantic progressive (grammar (b)).

Thus even though Deo's original model derives the progression from (a) to (d), it has no place for language stages like Early Modern English or Present-Day English. Importantly, this is not just a quirk caused by poorly selected A and Q: there is a principled reason why the problem appears for any reasonable A and Q as long as the population is infinite. Recall that according to payoff matrix A, each grammar i is most efficient in a population consisting of only i speakers. Note that this property should hold of any

14 The modeling problem Deo addresses with that boost is that the system does not progress to an all-(d) stage. With slightly different choices for the "independent" rates in Q, there won't be a high-(c)-share equilibrium and the system will progress to all-(d), cf. trajectories in SM 1, so no modification of the dynamics would be needed. However, such treatment does not solve the problem of intermediate grammars that this note addresses.

linguistically-reasonable A: we can hardly assume that speaking a different language from anyone else in the population will help you getting your meaning across! But then once the population gets close to everyone sharing the same language, in the deterministic dynamics implied by the infinite population, the speaker community will never move away from that highly communicatively efficient state. The constant rate of misacquisition will lead some "linguistic children" to adopt a novel grammar, but communicative efficiency will keep their share at some equilibrium value, as they will be much less successful communicators and thus won't provide attractive target grammars.[15]

But identifying the problem is already half the solution. As the problem is caused by the system's deterministic behavior stemming from the infinite population assumption, adopting the more realistic finite-population setting will help us solve it, as we show in the next section. The crucial difference in the finite setting is that sometimes, the system may temporarily jump into a less communicatively efficient state, and through such jumps, it can eventually get from one nearly-uniform population state to another.

3 Improving upon Deo's results: finite populations

Deo's infinite population develops deterministically, in accordance with the replicator-mutator evolutionary law in (1)/(2), and under the forces of communicative efficiency in A, playing the role of fitness relevant for natural selection, and misacquisition in Q, playing the role of mutation. In a finite population, the process becomes probabilistic rather than deterministic, which is in fact the key to improving on Deo's original results.

A finite population is always subject to random fluctuations. In population-genetics parlance, it is said that finite populations experience one more force, in addition to fitness and mutation: *genetic drift*, which is really just the name for the randomness in the finite-population process. That randomness allows the evolutionary process to depart from its expected diachronic trajectory. For example, consider the rate of misacquisition from target (c) to actually acquired (d) in Deo's Q, which is 3%. In an infinite population, the actual share of speakers misacquiring (d) instead of target (c) will always be exactly

15 For a more detailed explanation, see Yanovich 2016: Sec. 4. For illustration, cf. SM 1, with trajectories derived with different rates in Q, and SM 2, with Deo's original A and Q, but scaled differently by α and β. In no case do we see intermediate grammars getting close to 100% before giving way to the next stage in SM 1 and 2. Nor can we: see again the explanation.

3%. But in a finite population, that share will be a random number: with 3% as the probability for such misacquisition by an individual speaker, in some generation, it may so happen that a fewer number of speakers than expected misacquire their grammar that way, and in another generation, it could be a greater number. Crucially, the smaller the population, the easier it is to deviate farther from the expectation. For instance, with population size $N = 100$ where everyone speaks grammar (c), there is a 0.09% chance that 10% instead of 3% will misacquire (d). With $N = 1000$, that chance is only 6×10^{-25}. Genetic drift is stronger in smaller populations.

Such random deviations from the expected trajectory is what is crucial for our problem. Recall the explanation of the problem above. If the infinite population gets close to everyone sharing the same grammar, it becomes very communicatively inefficient to switch to another grammar. The communicatively optimal path of development, despite constant misacquisition, is for the population to maintain that nearly-universally shared grammar. But if the actual trajectory can probabilistically deviate from the optimal, mathematically expected path, then our population may at some point happen to venture so far from the near-uniform state that it will be possible for the next grammar in the cycle to get prominence. In other words, finite populations experience random fluctuations, which in turn may allow them to proceed from one communicatively very convenient state to another.

Here is how fitness, mutation and drift work for Deo's model with a finite population. Pure or near-pure states with dominance for intermediate grammars (b) and (c) are favored by communicative efficiency/fitness.[16] Directional movement $(a) \rightarrow (b) \rightarrow (c) \rightarrow (d)$ is favored by mutation matrix Q: it generally does not favor back-shifts to earlier stages, as rates of misacquisition going forward in the cycle are set to be higher than those going backward. Finally, jumps from one favored state to another are helped by drift, but this force is not sensitive to directionality: it is only mutation and fitness that distinguish between the grammars, while drift is blind to that.

These general conditions lead us to two types of reasonable evolutionary trajectories in Deo's model. Figure 2 presents the first type, arising under strong selection (=strong role for communicative efficiency) and strong drift (=small population size, relative to the strength of fitness and mutation). Strong selection makes the intermediate grammars really attractive. Moreover, grammar (c) with an obligatory progressive is more attractive than (b),

16 They are also actually favored by genetic drift, but explaining why is beyond the scope of this note.

as (c) is the only one that allows for perfect information transfer. Under such conditions, we see a time-limited stage of (b) dominance, followed by an extremely stable stage where almost everyone speaks (c). In this type of outcome, the process will eventually move along to (d), but it can take an enormous time for that to happen. This is so because in terms of communicative efficiency, (c) is much better than (d) as long as the cost k for having two forms in the grammar is low. So the fitness advantage of (c) makes it really hard for (d) to get hold even with relatively strong drift. It is worth noting that empirically, we do not actually know whether (d), where the old imperfective is lost, replaces (c) because of the forces internal to the progressive-imperfective cycle, or whether perhaps the loss of the old imperfective is an independent process of old-inflection loss that by necessity affects the progressive-imperfective grammar, but is not caused by anything internal to it.

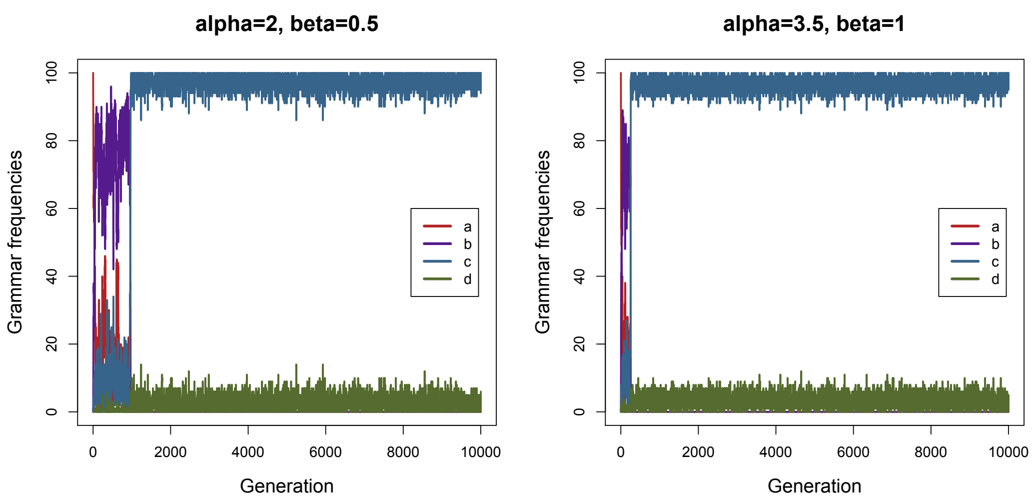

Figure 2 Strong drift, fitness relatively stronger than mutation. Left: $N = 100$, $\alpha = 2$, $\beta = 0.5$. Right: $N = 100$, $\alpha = 3.5$, $\beta = 1$. Note that with the population size constant, it is the *relative* strength of communicative efficiency, whose effect is scaled by α, and misacquisition, scaled by β, that matters.

The second reasonable type of outcome is given in Figure 3. Here, drift is made stronger than both mutation and fitness. (In practical terms, we increase the strength of drift by scaling down the effect of both communicative efficiency and misacquisition, via low α and β. Alternatively, one could decrease population size N, with similar effects.) A drift-dominated process is characterized by both a strong preference for pure states and a tendency to jump around between those pure states. Trajectories from different runs with the very same parameters can thus be very different, as we can see in the provided plots. With mutation giving the cycle its directional push, we get a nice progression of $(a) \to (b) \to (c) \to (d)$ in many instances of the process, but then we also see sometimes that the (b) stage is virtually skipped altogether (see the lower right plot in Figure 3), and other times, back-shifts to preceding stages of the cycle occur (see how in the lower left plot in Figure 3, the process gets close to skipping directly from (b) to (d), without passing through stage (c)). What we currently know of the empirical cycle does not rule out this type of outcome. For example, we lack fine-grained descriptions of the progressive-imperfective cycle that would have detected back-shifts and skipping of stages had they occurred!

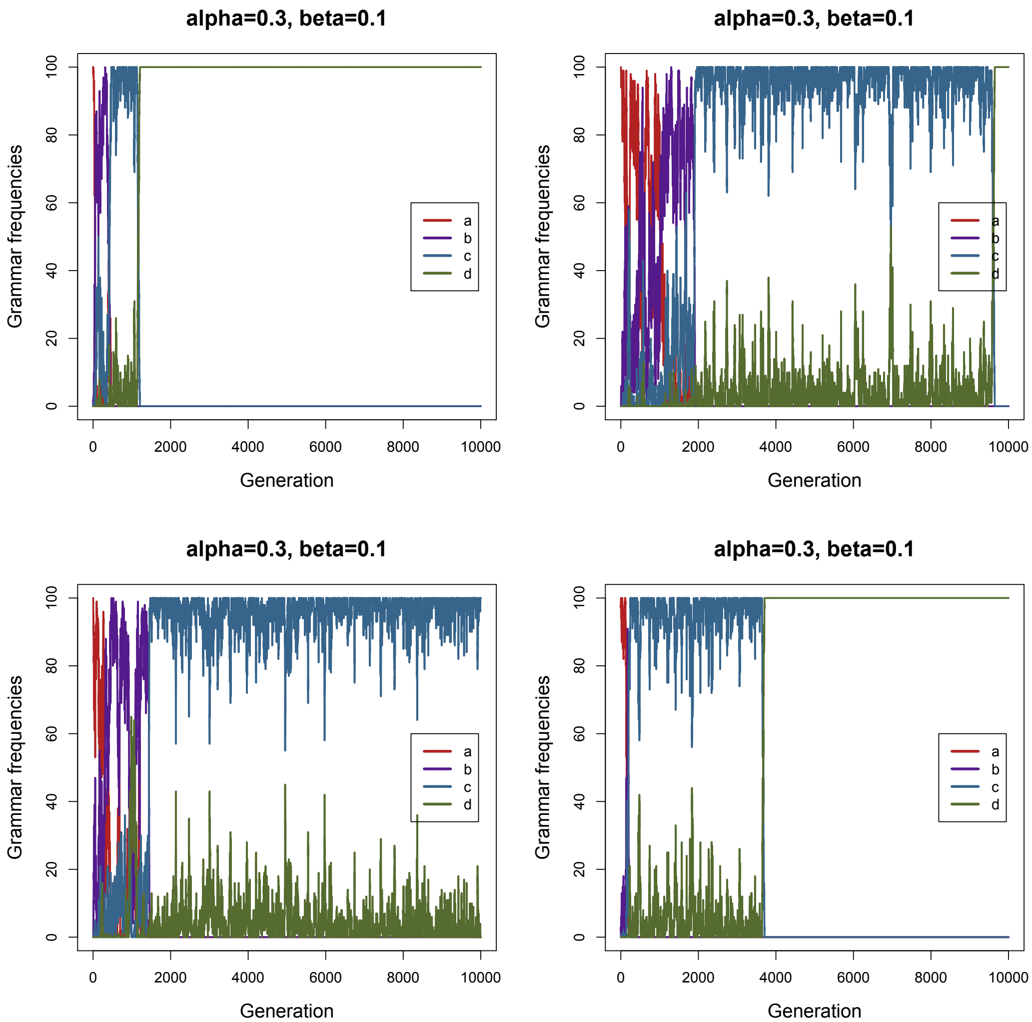

Figure 3 Very strong drift; mutation relatively stronger than fitness. All plots: $N = 100$, $\alpha = 0.3$, $\beta = 0.1$.

Let's sum up. First, we have solved the problem of Deo's original predictions. Trajectories in both Figure 2 and 3 show the population almost universally sharing both (b) and (c) grammars for considerable periods of time. This agrees with our empirical observations about the cycle. Second, our trajectories make specific, and not *a priori* expected, predictions about the fine-grained shape of the instantiations of the cycle. Under the type in Figure 2, we expect stage (c) to be able to hold for a very long time as long as there isn't an external push for losing the old imperfective. Under that in Figure 3, we do not expect to necessarily have an external push for moving from stage (c) to stage (d), but "in return" we expect to sometimes see stages almost skipped, especially the less communicatively efficient (b), and in rare cases, we expect returns back to a previous stage. It is worth noting that this two types of reasonable outcomes form a continuum: for example, under many parameter settings, stage (c) is relatively stable (as it is in the first type of outcome), even when the process can produce back-shifts and stage skipping (as in the second type). This stems from the fact that in Deo's game-theoretic analysis of progressive-imperfective communication, grammar (c) is the most efficient: it is the only one that allows for perfect communication, hence the mean communicative efficiency of a population where almost everyone speaks (c) is the highest, as long as the cost k of maintaining two forms X and Y does not become too high.

Whether we indeed see such types of trajectories should be checked empirically. Importantly, before the predictions were generated, it would not be obvious why the fine-grained behavior of the actual progressive-imperfective cycles needed to be studied. Just as spelled-out theories of grammar make predictions for synchronic observations that can either confirm or refute the theories, evolutionary-modeling theories can similarly generate predictions capable of guiding empirical investigations. Three particularly important questions are: (i) how fast can languages develop an obligatory-progressive system, that is, grammar (c), out of a system with only one generalized form for both progressive and imperfective meanings? Can they basically skip stage (b) with an optional new progressive? (ii) Can languages experience back-shifts in the cycle, reverting to an earlier stage? (iii) For the shift from (c) and (d), wherein the old imperfective form is lost, can we tell if that shift is caused by factors internal to the progressive-imperfective system, or rather by some external force making the old form to retire?

Note that in addition to the reasonable outcomes, we also see rather implausible trajectories arising under some sets of evolutionary parameters. The two reasonable types of predictions above are thus not robust. So the ultimate fate of the model crucially depends on whether we can eventually find evidence that the real-life evolutionary parameters fall into the range that is needed for reasonable predictions to arise. Consider two features in particular: the number of generations needed, and the small population sizes. For the number of generations, it is obvious that if we interpret generations as biological generations, then requiring several thousands of them for the cycle to proceed is a bad prediction. But "model generations" may also be interpreted in terms of grammar switches in adulthood (see, for instance, Raumolin-Brunberg 2009), cf. fn. 13. Then the number will not necessarily be a problem. As for the population size, the simple model we used is actually somewhat misleading. It assumes that everybody in the population is equally likely to provide a target grammar for every new speaker. But in real populations, especially large ones, speakers are organized in complex social networks of interaction, and a specific child does not have equal chances of picking any adult in the population at large as the model. Structure in populations isolates speakers from each other, increasing genetic drift. Whether this effect is enough to produce the cycle's trajectories in modeled communities of realistic size is an open question, but at least there is no reason to reject the current account as *a priori* implausible.

Many open questions remain, and much new formal research needs to be done before our evolutionary models of semantic change mature into well-behaved, well-understood, and obviously useful tools of diachronic semantic inquiry. But even today we can see such models making new testable empirical predictions, as shown above. In synchronic formal semantics, the tools of higher-order logic are indispensable, but it was not easy to realize that fact, those tools have steep learning curves, and it took several decades for the analytical apparatus of modern formal semantics to mature. Evolutionary modeling of semantic change is still in its early and exciting stages, but it can eventually develop into a similarly indispensable instrument for doing diachronic formal semantics.

Supplementary materials

Supplementary materials are available at https://doi.org/10.3765/sp.10.17s.

SM 1. Infinite-population trajectories in the Imperfective Game, under a number of combinations of four independent misacquisition rates in matrix Q. The figures illustrate both the dependence of the trajectories on the exact form of Q, and the limits on how much one can vary the trajectories by changing Q. A as in Deo 2015, that is as in Table 2 with $k = 0.01$.

SM 2. Infinite-population trajectories illustrating the effect of varying α and β. A and Q as in Deo 2015, that is as in Table 2 with $k = 0.01$ and in Table 3, respectively.

References

Burger, J., M. Kirchner, B. Bramanti, W. Haak & M. G. Thomas. 2007. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proceedings of the National Academy of Sciences* 104(10). 3736–3741. https://doi.org/10.1073/pnas.0607187104.

Deo, Ashwini. 2015. The semantic and pragmatic underpinnings of grammaticalization paths: The progressive to imperfective shift. *Semantics and Pragmatics* 8(14). 1–52. https://doi.org/10.3765/sp.8.14.

Der, Ricky, Charles L. Epstein & Joshua B. Plotkin. 2011. Generalized population models and the nature of genetic drift. *Theoretical Population Biology* 80(2). 80–99. https://doi.org/10.1016/j.tpb.2011.06.004.

Eckardt, Regine. 2006. *Meaning change in grammaticalization: An enquiry into semantic reanalysis*. Oxford: Oxford University Press.

Gillespie, John H. 2004. *Population genetics: A concise guide*. 2nd edn. Baltimore: Johns Hopkins University Press.

Jones, Bryony L., Tamiru O. Raga, Anke Liebert, Pawel Zmarz, Endashaw Bekele, E. Thomas Danielsen, Anders Krüger Olsen, Neil Bradman, Jesper T. Troelsen & Dallas M. Swallow. 2013. Diversity of lactase persistence alleles in Ethiopia: Signature of a soft selective sweep. *American Journal of Human Genetics* 93(3). 538–544. https://doi.org/10.1016/j.ajhg.2013.07.008.

Niyogi, Partha. 2006. *The computational nature of language learning and evolution*. Cambridge, MA: MIT Press.

R Core Team. 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Raumolin-Brunberg, Helena. 2009. Lifespan changes in the language of three early modern gentlemen. In Arja Nurmi, Minna Nevala & Minna Palander-Collin (eds.), *The language of daily life in England (1450-1800)*, 165–196. Amsterdam: Benjamins. https://doi.org/10.1075/pbns.183.10rau.

Sandholm, William H. 2010. *Population games and evolutionary dynamics*. Cambridge, MA: MIT Press.

Swallow, Dallas M. 2003. Genetics of lactase persistence and lactose intolerance. *Annual Review of Genetics* 37. 197–219. https://doi.org/10.1146/annurev.genet.37.110801.143820.

Yanovich, Igor. 2016. *Analyzing imperfective games: The companion paper*. Tech. rep. Universität Tübingen & Carnegie Mellon University. http://ling.auf.net/lingbuzz/002652.

Igor Yanovich
DFG Center for Advanced Studies "Words, Bones, Genes, Tools"
Universität Tübingen
Rümelinstraße 23, Rm. 604b
Tübingen, 72076, Germany
igor.yanovich@uni-tuebingen.de