Working Papers

Institute of Mathematical Economics

420

August 2009

Voronoi Languages

Equilibria in Cheap-Talk Games with High-Dimensional Types and Few Signals

Gerhard Jäger, Lars Koch-Metzger and Frank Riedel

IMW · Bielefeld University
Postfach 100131
33501 Bielefeld · Germany

email: imw@wiwi.uni-bielefeld.de
http://www.wiwi.uni-bielefeld.de/˜imw/Papers/showpaper.php?420
ISSN: 0931-6558

Voronoi Languages[†]

Equilibria in Cheap-Talk Games with High–Dimensional Types and Few Signals

By Gerhard Jäger[*], Lars Koch-Metzger[**] and Frank Riedel[**]

Tübingen University[*] *and Bielefeld University*[**]

— this Version: August 13, 2009

Abstract

We study a communication game of common interest in which the sender observes one of infinite types and sends one of finite messages which is interpreted by the receiver. In equilibrium there is no full separation but types are clustered into convex categories. We give a full characterization of the strict Nash equilibria of this game by representing these categories by *Voronoi languages*. As the strategy set is infinite static stability concepts for finite games such as ESS are no longer sufficient for Lyapunov stability in the replicator dynamics. We give examples of unstable strict Nash equilibria and stable inefficient Voronoi Languages. We derive efficient Voronoi languages with a large number of categories and numerically illustrate stability of some Voronoi languages with large message spaces and non-uniformly distributed types.

[†] We would like to thank Joseph Hofbauer from University of Vienna and the session auditory of the North American Summer Meeting of the Econometric Society in Boston 2009 for their comments. The research that led to this paper has been supported by project A6 of the SFB 673 "Alignment in Communication" of the German Research Foundation, which is gratefully acknowledged.

JEL subject classification. C72, C73, D82, D83

Key words. Cheap Talk, Signaling Game, Communication Game, Dynamic stability, Voronoi tesselation

1 Introduction

In many situations where one person signals information to another, the complexity of the sender's information is much higher than the variety of possible signals. This is so in the basic act of speaking when an agent tries to transmit the information of a sensation to a hearer. The sensation is a high–dimensional object containing shape, size, color, temperature etc., but the hearer understands only finitely many words. The speaker thus has to aggregate a lot of possible types under one name. Another simple example is the way a baseball cap is worn (which serves as a social signal among certain sub-cultures). There are only finitely many easily distinguishable ways how to wear a baseball cap: the brim facing forward, backward, to the left or to the right. These signals may convey certain information about the type of the wearer of the cap, like age, group membership, musical taste etc. So the space of possible types is virtually unbounded, while the space of signals is small. In the famous job market signaling game of Spence (1973), skills – albeit frequently modeled as $0 - 1$ or one–dimensional — can be naturally thought of as multidimensional — from verbal ability over mathematical skills to social competence a wide array of properties describe a worker. On the other hand, workers have usually only finitely many different education levels (high school, college, university) to choose from. In finance, rating agencies use a discrete grid to signal information about the credit quality of firms while the underlying information is certainly much more complex.

In all these situations, we have a signaling game in which perfect separation is impossible as there are too few signals. In fact, the type space is much bigger than the signal set. One way to model such a situation formally is to assume that types come from a continuum, a convex subset of n–dimensional Euclidean space, e.g., whereas signals come from a finite alphabet. We study these games here and perform static as well as dynamic evolutionary analysis of its equilibria. To keep things simple, we assume that the interests of sender and receiver are identical. So payoff would be maximal if the receiver would always correctly guess the sender's type. However, such perfectly separating equilibria are clearly not possible.

We show that strict Nash equilibria of our game are given by what we call *Voronoi languages*[1]. The sender partitions the type set into convex sets;

[1] We use a language that is inspired by linguistics throughout. Nevertheless, our "Voronoi languages" have natural interpretations in the job market or other contexts.

to each signal, there corresponds one such set, or cell. We show that the cells form a so–called Voronoi tesselation of the type space: for every signal, there is a certain prototype. One can think of that prototype as the "typical" representative of the class, as the "typical" shade of blue, or the "typical" professor of economics, the "typical" politician and so on. Upon seeing her type, the speaker chooses the prototype that is closest to that point in the type space and transmits the corresponding signal to the receiver. The prototypes induce then a partition of the type space into convex polyhedra; such partitions are called Voronoi tesselations.[2] There are of course a plethora of possible Voronoi tesselations, but only few of them form part of Nash equilibria. For such an equilibrium, the prototype must also be the best possible interpretation for the receiver. The receiver — knowing that the signal used corresponds to a certain subset of types — chooses the interpretation that leads to the minimal expected loss. In statistical terms, he chooses the best conditional estimate for that type set. A Voronoi language thus consists of a Voronoi tesselation where the prototypes are also the best Bayesian estimates.

As we show by example, this usually leaves very few equilibria (up to the obvious inessential multiplicities). On the unit interval, when types are uniformly distributed and similarity is measured by the usual distance, there is only one Voronoi language. In the unit square, again with uniform types and Euclidean distance, there are two Voronoi languages when there are two words. The first (and better) one, separates the square into left and right. The second one uses the diagonal to partition the square. This example also shows that not all Voronoi languages (and strict Nash equilibria) are efficient. When we use the Euclidean distance, an efficient language has to minimize the sum of conditional variances of the errors. Partitioning the square into two rectangles leads to a smaller variance than partitioning into two triangles.

We then go on and study the evolution of such signaling structures. As is by now well known in such games with a continuum of strategies, strict Nash equilibria need not be dynamically stable.[3] Indeed, in our above example,

[2] Voronoi tesselations appear naturally in other disciplines such as geography (basins of drainage), data compression (where they are used in vector quantization) or climatology (where they are referred to as Thiessen polygons).

[3] Oechssler and Riedel (2002) show that this is not an artefact of the continuum model. When a strict Nash equilibrium is unstable in the continuum, this means that the basin of attraction of a strict Nash equilibrium vanishes as the grid size becomes arbitrary small.

the diagonal language is not stable under replicator and similar dynamics, and evolution thus converges to the efficient language.

This begs the question if evolution always leads to efficient languages. On the one hand, we show that efficient languages are stable: once an optimal language has been found, no mutants can invade. This is an easy consequence of the fact that the payoff function in this common interest game is a Lyapunov function.

On the other hand, evolution can also lead to inefficient languages. To this end, we consider a rectangle in dimension 2 with two different side lengthes; with two words and uniformly distributed types, there are two natural stable Voronoi languages. The first one partitions the rectangle into "up" and "down", and the second language into "left" and "right". Both languages are local minima of the Lyapunov function (average payoff), and thus stable. Only one of them is efficient, though. This suggests that evolution does not necessarily find optimal languages[4].

Finally, we develop a numerical algorithm that allows to find Voronoi languages. This is important when the number of words is large or the type distribution is not uniform because it is then usually impossible to find the equilibria in explicit form. We use this algorithm to show that partitioning the square into squares is a stable Voronoi language for small alphabets whereas it is not stable as the number of words grows. In this case, evolution tends to Voronoi tesselations that look like bee hives, consisting of regular hexagons. We also provide illuminating examples for non–uniform distributions. In these cases, languages tend to distinguish very sharply types that have high frequency whereas big regions are used for one word when the types are not very frequent.

It would lead us too far to review the vast literature on signaling games. The seminal paper on cheap talk games as we study them here is Crawford and Sobel (1982). These authors focus on strategic issues created by slightly misaligned interests. Many papers[5] investigate issues concerning efficiency and cooperation in cheap talk games. These papers assume that there are at least as many signals as types or that the utility is either 1 (success) or 0

[4] It might well be possible that stronger concepts like Schlag's evolutionarily absorbing set or Matsui's cyclycally stable set yield stronger conclusions. As these concepts have not yet been extended to games with a continuum of strategies, we leave this question for further research.

[5] Robson (1990), Matsui (1991), Schlag (1993), Sobel (1993), Blume, Kim, and Sobel (1993), Wärneryd (1993), or Trapa and Nowak (2000) are prominent examples.

(failure). To the best of our knowledge, our game with a multidimensional type space and payoffs depending on the distance of types and interpretations has not been studied before[6].

Azrieli and Lehrer (2007) axiomatically derive convex categories for models with at least two dimensional type spaces. For their approach, extended prototypes as justified by Gardenfors (2000) are necessary. They capture the 'size' of a category, which is captured by an explicit measure on the type space in our model. Although our model is described in language terms, it can be well applied to alternative settings. For example, Azrieli (2009) applies convex categories to a model of political election. In a model similar to ours Fryer and Jackson (2008) also address the question of efficient categorization. Considering uniformly distributed continuous types, the authors focus on binary realizations and do not further investigate stability issues. A recent experiment on announcement games by Agranov and Schotter (2008) indicates that coordination on a certain language seems more achievable if few words are available.

The paper is set up as follows. Section 2 develops the game we consider. Section 3 studies efficient languages. We characterize strict Nash equilibria in Section 4. Section 5 contains our dynamic evolutionary analysis, and Section 6 provides the numerical algorithm and examples. Section 7 concludes.

2 Model and Notation

The sender has a type $t \in T$, where T is a convex and compact subset of $\mathbb{R}^L$ for some $L \geq 1$ that has nonempty interior. He chooses a word (signal) $w \in W := \{w_1, \ldots, w_N\}$ from a finite language and sends it to the receiver. The receiver interprets w as some point $i \in T$. Both players want type t and interpretation i to be as similar as possible. We assume that $l\left(\|t-i\|\right)$ measures the loss of the players where the function $l : \mathbb{R}_+ \to \mathbb{R}$ is convex and strictly increasing.[7] A natural choice that we consider frequently below is the square Euclidean distance $\|i-t\|^2$. The probability of types is described by

[6] A setting similar to ours was proposed in Jäger and van Rooij (2007) and worked out in some detail in Jäger (2007). We generalize the model to continuous type spaces and provide the full game–theoretic analysis. The evolutionary analysis uses the tools developed in Oechssler and Riedel (2001), Oechssler and Riedel (2002), and Cressman, Hofbauer, and Riedel (2006).

[7] For one dimensional type space we require $l(\cdot)$ to be strictly convex.

an atomless distribution F on T with strictly positive and continuous density $f : T \to \mathbb{R}_+$.

A (pure) strategy for the sender is a measurable function $w : T \to W$. We denote by Σ the set of all sender strategies. A (pure) strategy for the receiver is a vector $i = (i_1, \ldots, i_N) \in T^N$ where i_j denotes the interpretation of the word w_j. The expected loss of players is then

$$L(w, i) = \int_T l\left(\|t - i_{w(t)}\|\right) F(dt)\,.$$

Note that null sets play no role for the expected loss. Hence, we ignore them in the sequel when we characterize strategies.

3 Efficient Languages

To start with, we study what the two players can achieve in cooperation. Ideally, we might think of super–rational players who have a meta–language to communicate with each other; before playing, they meet in an ideal place to discuss their efficient strategy. Formally, we call a language (w, i) efficient if it minimizes the loss $L(w, i) = \int_T l\left(\|t - i_{w(t)}\|\right) F(dt)$.

Before coming to the proofs, let us give a short synopsis of the results. If there were as many words as types, the players would clearly choose a language that distinguishes perfectly all private information (a fully separating equilibrium in the language of game theory). In our situation, this is not feasible as the type space is a continuum and the set of words is finite. Nevertheless, efficient languages are “as separating as possible”, i.e. they use all available words and attach different meanings to them. Suppose word w_n were unused in an efficient language. The sender could then split up the set of types which lead to word w_1 into two convex sets, one of which serving for w_1 and the other serving for w_n. By choosing appropriate interpretations, the resulting language has lower expected loss. Hence a language with unused words cannot be efficient. The sender will thus choose a partition $(C_k)_{k=1,\ldots,n}$ of the space T and say word w_k whenever his type t is in the cell or *category* C_k.

Given that the receiver uses prototypes $i_k \in T$, we can ask what the optimal partition is. The sender wants the prototype to be as close to his type as possible. Hence, he will say w_k whenever the prototype i_k is closest to his type t among all prototypes. Such a partition of the type space is called a *Voronoi tessellation* of the space. At first glance, it might seem that we cannot say much more. This is not true, however. Given such a partition, the receiver in turn has to choose a "prototype" $i_k \in T$ for each word w_k that describes the average type in C_k optimally (given the environment or prior F). An optimal interpretation consists thus of Bayesian estimators for each cell C_k.

Summing up, efficient languages consist of what we call *Voronoi languages with full vocabulary*—a Voronoi tessellation of the space T that is induced by points i_k which are at the same time Bayesian estimators for the average type in each cell. Depending on the reader's intuition, you might expect to find a plethora or very few of such efficient languages. We illustrate by examples that there are usually very few Voronoi languages (up to the obvious symmetries, of course).[8]

Let us come to the formal analysis.

Definition 1 *A language (w, i) consists of a measurable mapping $w : T \to W$ (the signaling strategy) and points $i \in T^N$ (the interpretation). A language (w, i) has full vocabulary if* $\operatorname{range} w = W$.

We show now that our problem is well–posed, i.e. that efficient languages (w, i) which minimize $L(w, i)$ exist. A slight technical problem comes from the fact that the payoff $l\left(\|t - i_{w(t)}\|\right)$ is not continuous in t, in general. We proceed as follows. We first show that one can restrict attention to strategies w that are induced by Voronoi tessellations. As Voronoi tessellations can be described by their center points $i_k \in T$, we can study now an auxiliary payoff function which only depends on N points in T. As T is compact, it is enough to show continuity of this function. This is done by noting that $l\left(\|t - i_{w(t)}\|\right)$ jumps only at the boundaries of Voronoi cells which form a Lebesgue null set. Hence, the auxiliary payoff function is continuous by Lebesgue's theorem. As a consequence,

Lemma 1 *Efficient languages exist.*

The proof of this and all other results can be found in the appendix.

We turn now to an analysis of efficient languages. Let us begin with a detour. So far, we have not even discussed the possibility of mixing, or

[8] Crawford and Sobel (1982) call a class of such symmetric equilibria "essentially equivalent".

randomized strategies (and we will not do so later on). For one paragraph, we will allow for this possibility — just to show that mixing is not efficient. This is quite plausible: the players have no reason to introduce randomness in their communication when they cooperate.[9] A mixed strategy for the sender is a measurable mapping $\omega : T \to \Delta W$ where ΔW denotes the set of probability vectors over W. We denote by $\omega_k(t)$ the probability that the sender chooses word w_k if in type t. A mixed strategy for the receiver consists of probability measures $(\mu_k)_{k=1,\ldots,N}$ over T.[10] The generalized loss function for such strategies is then

$$L(\omega, \mu) = \int_T \sum_{k=1}^{N} \int_T l\left(\|t - i\|\right) \mu_k(di) \omega_k(t) F(dt) \, .$$

Lemma 2 *For every language (ω, μ) in non-degenerate randomized strategies, there is a pure strategy language (w, i) which is strictly better.*

From now on, we thus return to pure strategies (w, i).

Our next rather obvious point is that players should use all available words given that there is no cost in using them. The proof uses the fact that F is atomless. When a language does not use one word w_N, say, one can split a used word, w_1, say, in two words, and obtain a better language. It is also clear that the receiver should interpret different words differently (as they represent different convex areas of the type space T with pairwise disjoint interiors).

Lemma 3 *Efficient languages (w^*, i^*) have full vocabulary and interpretations i_k^* are pairwise distinct.*

We can thus focus on languages in which all interpretations are pairwise distinct. Given that the receiver uses the pairwise distinct points (i_k), what words should the sender choose if in type t? Clearly the word that leads to the interpretation i_k which is closest to t among all interpretations.

[9] There are, of course, mixed, or partially mixed Nash equilibria, and mixing can be a best reply. The convexity of the loss function induces risk aversion for the players. Their payoff is thus not increased by mixing.

[10] Whenever we speak of measurability, probability etc. we think of T as endowed with the Borel σ–field.

Lemma 4 *In efficient languages (w^*, i^*), the sender uses a Voronoi tessellation corresponding to i^*, i.e. F–almost everywhere*

$$w^*(t) = \operatorname{argmin}_{j=1,\ldots,N} \|t - i_j^*\| \,. \tag{1}$$

Note that the above strategy is not uniquely defined at points t that have equal distance to two or more interpretations. As these points form a null set, we can ignore this ambiguity; without loss of generality, we always take the word with smallest index in this case.

It is quite easy to see (cf. for instance Okabe, Boots, and Sugihara (1992) for a proof) that in Euclidean spaces, the interior of each cell of a Voronoi tessellation is a convex set. (To see why, please observe that for each pair of prototypes x and y, the set of points that is closer to x than to y forms an L-dimensional half-space that is bounded by the hyperplane of points that are equidistant to x and y. A half-space is evidently a convex set. A Voronoi cell is an intersection of finitely many half-spaces, and the intersection of convex sets must be convex again.) So we have the

Corollary 1 *In efficient languages (w^*, i^*), the sender uses convex categories, i.e. for each i_j^*, $w^{*-1}(i_j^*)$ is (up to a null set) a convex set, the intersection of a convex polyhedron with the type space T.*

Let us now come to the receiver. Given that the sender uses a Voronoi tessellation of which each cell has positive measure, the receiver has to determine an optimal interpretation. By Bayes' rule, she has to choose an optimal estimator given that she knows the type to be in that cell.

Definition 2 *Let $C \subset T$ be a convex set with positive measure. Call*

$$b(C) = \operatorname{argmin}_{i \in C} \int_C l\left(\|t - i\|\right) F(dt)$$

the Bayesian estimator conditional on C.

Remark 1 *Note that the Bayesian estimator is uniquely determined. This follows from Jensen's inequality. We get the strict inequality because the integrand $l\left(\|t - i\|\right)$ is convex, increasing, and not linear in i.*

Let us state the best estimators for the quadratic and linear loss function.

Example 1 *1. For $l(d) = d^2$, the best estimate is the conditional expectation,*

$$b(C) = \mathbb{E}[t|t \in C] := \frac{1}{F(C)} \int_C t\, F(dt)\,.$$

2. For $l(d) = d$ and $L \geq 2$, the best estimator is the (generalized) conditional median type given the cell C.

Lemma 5 *In efficient languages (w^*, i^*), the receiver uses the best interpretation of the partition induced by w^*, i.e.*

$$i_k = b(C_k^*)$$

for

$$C_k^* = \{t \in T : w^*(t) = w_k\}\,.$$

We summarize our findings in a definition.

Definition 3 (Voronoi Language) *A Voronoi language (w, i) consists of a Voronoi tessellation for the sender and an Bayesian estimator interpretation for the receiver. i.e. we have both*

$$w^*(t) = \mathrm{argmin}_{j=1,\ldots,N} \|t - i_j^*\| \ F - a.s. \tag{2}$$

$$i_k = b(C_k^*) \qquad (\text{ for } C_k^* = \{t \in T : w^*(t) = w_k\})\,. \tag{3}$$

Any language with an optimal sender strategy induces a Voronoi tesselation. We additionally assume that a Voronoi language satisfies receiver optimality. The new concept allows us to describe efficient languages succinctly.

Theorem 1 *Efficient languages are Voronoi languages with full vocabulary.*

To get a better intuition, we start with two (highly idealized and simple) examples where there are only two words and types are uniformly distributed. On the unit interval $[0, 1]$, there is only one Voronoi language with full vocabulary (which is also the unique efficient language). On the unit square, there are two Voronoi languages with full vocabulary (up to symmetries). Only one of them is efficient. The converse of the above theorem is thus not valid.

Example 2 *Consider the unit interval $T = [0,1]$ with the uniform distribution $F(x) = x$, quadratic loss $l(d) = d^2$, and two words $W = \{w_1, w_2\}$. The two words have the obvious everyday meaning of "left" and "right". The efficient Voronoi language has $w^*(t) = w_1$ for $t \leq 1/2$ and $w(t) = w_2$ else. The best interpretation is $i_1^* = 1/4, i_2^* = 3/4$. Let us quickly show that this is the only Voronoi language*[11] *with full vocabulary here. If we denote by K the threshold that separates the two Voronoi cells, we must have*

$$i_1 = K/2$$
$$i_2 = (1+K)/2$$

as Bayesian estimators and

$$K = (i_1 + i_2)/2$$

because K must correspond to the Voronoi tessellation induced by i_1 and i_2. Substituting i_1, i_2 in the third equation, we get $K = 1/2(1/2+K)$, or $K = 1/2$, and then $i_1 = 1/4$ and $i_2 = 3/4$ as desired.

In the previous example, efficient and Voronoi languages coincide. This need not be the case, as we now illustrate.

Example 3 (Not all Voronoi languages are efficient) *Consider the unit square $[0,1]^2$, with the uniform distribution, quadratic loss $l(d) = d^2$, and two words $W = \{w_1, w_2\}$. A typical Voronoi tessellation consists here of two points that lead to two trapezoids as illustrated in Figure 1. The two border cases are the* horizontal *and* diagonal *tessellation of Figure 2. One might guess that trapezoid tessellations that are symmetric around the center point $(0.5, 0.5)$ are Voronoi languages. This is not true, however because the center of gravity (the Bayesian estimator) of a trapezoid does not coincide with a point that generates the cell (see Figure 3). See the appendix for the geometric construction of centers of gravity. Solving a polynomial shows that the diagonal and the vertical language are the two unique Voronoi languages with full vocabulary. The diagonal is not efficient, however, because it leads to a loss*

$$2 \cdot \int_0^1 \int_0^{1-y} (x-1/3)^2 + (y-1/3)^2 dxdy = 1/9 \simeq 0.111,$$

[11] up to the obvious symmetry, of course

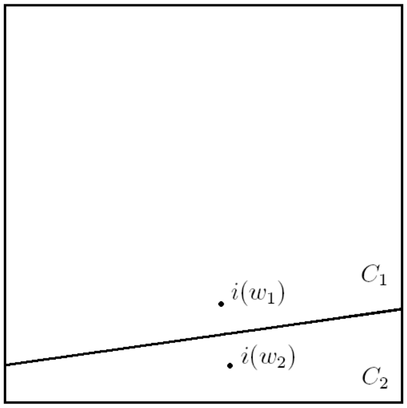


Figure 1: A Voronoi tesselation

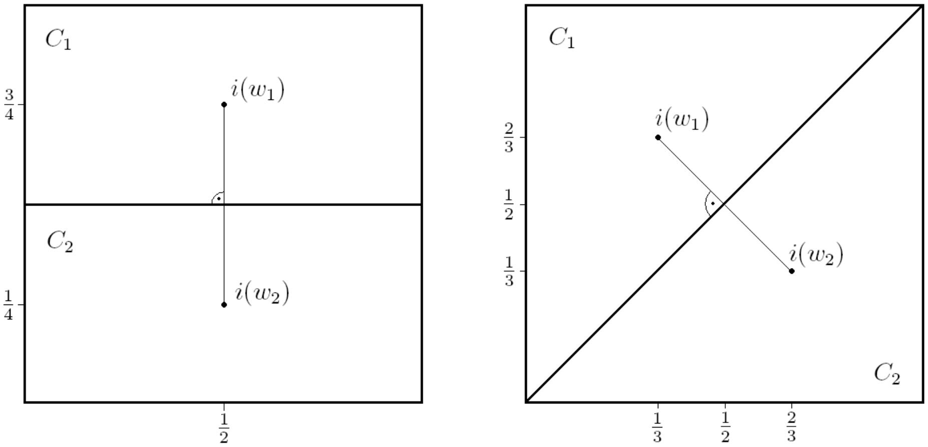


Figure 2: horizontal and diagonal Voronoi tesselation

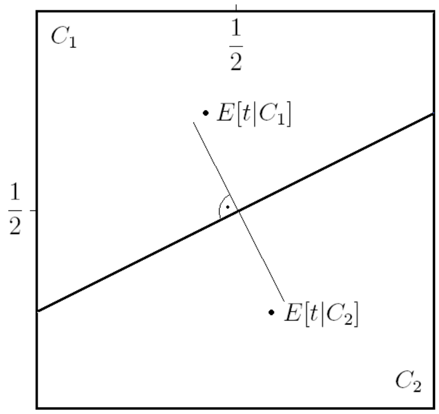


Figure 3: Centers of gravity do not always generate their Voronoi tesselation.

whereas the horizontal language has loss

$$2 \cdot \int_0^1 \int_0^{1/2} (x - 1/4)^2 + (y - 1/2)^2 dxdy = 5/48 \simeq 0.104\,.$$

To give some intuition, the sum of conditional variances is lower in the horizontal language than in the diagonal language.

4 Pure Strategy Nash Equilibria

Although cooperative solutions can be achieved in ideal situations where the players have the possibility to use a meta–language for before–play communication, the everyday situation is different. Here, we rather have to guess what our partner might mean with his words—not an easy task. This situation is better modeled as a noncooperative signaling game between the two players. Let us assume rationality of the players for the moment (we turn to the more realistic case of bounded rationality later on).

As in all signaling games, there is a plethora of Nash equilibria. We focus here on strict equilibria where the best replies of both players are unique.[12]

[12] Strictly speaking, we require only that the sender's best reply is $F-$almost surely unique (changes on a null subset of T do not influence the payoff). As the expected

In this case, we can make use of our optimality analysis of the preceding section.

First, let us note that strict Nash equilibria share with efficient languages the feature that all words are being used. The argument is different from the one we used for efficient languages, though. For efficiency, we use the fact that the average loss can be reduced by using more words. Such a cooperative argument does not work in the game–theoretic setting. Instead, we rely on $F-$ a.s. strict Nash equilibrium to exclude such a phenomenon. If the sender never uses a word, say w_N, then the receiver is indifferent between all interpretations for w_N, and the best reply is not unique.

As for efficient languages, the sender's best reply to a given interpretation is the corresponding Voronoi tessellation. Similarly, the receiver's best reply to a partition consists of the Bayesian estimator—here, the arguments are identical to those in Lemmata 4 and 5.

Conversely, every Voronoi language with full vocabulary consists of a pair of mutually best replies, and is thus a Nash equilibrium. The sender's Voronoi tessellation is the (almost sure) unique best reply (there is indifference at the points that are equidistant to two or more interpretations, a null set). The receiver's best reply is unique because of the strict convexity of the loss function $i \mapsto l(\|t-i\|)$, compare Remark 1.

Theorem 2 *Every Voronoi language with full vocabulary is a strict Nash equilibrium and vice versa.*

We thus have a full characterization of strict Nash equilibria. In particular, we see that inefficient languages can arise even if we impose the relatively strong condition of strictness on the set of Nash equilibria, compare Example 3 above. Rational communication does not necessarily result in efficient signaling systems.

losses are invariant with respect to null subsets (as for example the border between two categories), all strategies are optimal for these sets. Hence the notions of weak perfect Bayesian equilibrium and Nash equilibrium do coincide here. Note that there are sequential equilibria (Kreps and Wilson (1982)) that have alternative best replies as for example the one word language, the pooling equilibrium in which the sender always sends the same word and the receiver's interpretation is $E[t]$ for any word.

5 Evolution of Voronoi Languages

Our current language is not a fixed system, rather a fluent and flexible body of words and rules that is constantly evolving. As such, it is shaped by the typical forces of selection and mutation that govern evolution. We are thus led to study dynamical systems that describe possible evolutionary dynamics.

On the technical side, we face here a rather complicated dynamical system because a population is described by a probability measure over all strategies—and strategies are pairs of signaling systems, i.e. simple measurable functions on T with values in W and interpretations, points in T^N. For several dynamics, the technical foundations for the study of the replicator (Oechssler and Riedel (2001), Oechssler and Riedel (2002), Cressman, Hofbauer, and Riedel (2006)), payoff–monotone (Heifetz, Shannon, and Spiegel (2007)), and Brown–von–Neumann–Nash dynamics (Hofbauer, Oechssler, and Riedel (2009)) have been worked out. Although our strategy space is slightly more general than in some of the cited papers, the general results of these papers hold true in our setting.

For our dynamical considerations, we consider the symmetrized version of the game. Let us suppose that agents are equally often in the role of receiver and sender, and every agent thus chooses both a sender strategy v or $w \in \Sigma$ as well as a receiver strategy i resp. $j \in T^N$. Then the expected loss of an agent using language (v, i) and meeting an agent using language (w, j) is

$$\Lambda((v,i),(w,j)) = 1/2(L(v,j) + L(w,i)) .$$

A population of agents is described by a probability distribution $P(dw, di)$ over the strategy set $\Gamma := \Sigma \times T^N$ of the symmetrized game. For two such distributions P and Q, we can extend the symmetrized loss function in the usual way by setting

$$\Lambda(P,Q) = \int_\Gamma \int_\Gamma \Lambda((v,i),(w,j)P(dv,di)Q(dw,dj) .$$

The dynamic analysis is greatly simplified by the fact that average loss is decreasing along the paths of typical selection and innovative dynamics, as usual in common interest games.

Lemma 6 (Fundamental Law of Natural Selection) *The symmetrized payoff function is a Lyapunov function for the replicator (more generally, regular, payoff–monotone) and the Brown–von Neumann–Nash dynamics.*

Technically, it is important to show that the loss function is continuous with respect to the weak topology for probability measures because we can only expect convergence in the weak topology, in general[13].

Lemma 7 *The payoff function is continuous with respect to the weak topology.*

To prepare the dynamic analysis, we describe the relation of some static stability concepts to dynamic stability.

Let us have a brief look at games with finite strategy sets. In those games asymptotic stability with respect to payoff monotonic evolutionary dynamics implies ESS, evolutionary stability, see for example Ritzberger and Weibull (1995). On the other hand, asymptotic stability is implied only for two player games (here: sender, receiver) in the replicator dynamics. According to Maynard Smith (1974), a strategy (a say), is *evolutionary stable* (ESS) if there is an invasion barrier ϵ such that if a subgroup of the population with size $\eta \leq \epsilon$ does not fare better using any strategy (b say) in the sense that $\Lambda(a, (1-\eta)a+\eta b) < \Lambda(b, (1-\eta)a+\eta b)$.

The invasion barrier ϵ has a twofold implication:

i) on the one hand it allows for mutant strategies b that are arbitrary different from a if the subgroup of deviating agents is small, while

ii) on the other hand, a being ESS implies that an arbitrary large subgroup of agents using strategy a has lower expected losses against $\tilde{a}_\epsilon = (1-\epsilon)a + \epsilon b$ as long as the induced strategy $\tilde{a}_\epsilon$ is close enough to a. It is exactly the combination of these two properties that requires a careful consideration in games with a continuum of strategies. In such games a mixed strategy is a density function on a continuum. Now the meaning of 'strategy α is *close* to strategy β' crucially depends on the choice of topology because meanings *i*) and *ii*) do not generally coincide. The strong topology (or variational- or supremum norm) considers closeness in the sense of *i*), which is the property that Vickers and Cannings (1987) and Bomze and Pötscher (1989) (when defining *uninvadability*) consider to be the relevant property. By contrast Eshel (1983) (when defining *continuously stable strategies*, CSS) and Apaloo (1997) (when defining *neighborhood invader*

[13] See Oechssler and Riedel (2002) and Hofbauer, Oechssler, and Riedel (2009) for an extended discussion of this point.

strategy, NIS) consider requirement *ii)* . If one considers a strategy α to be close to β in the sense of *i) or ii)*, the right topology to apply is the topology of weak convergence (or Prohorov metric). Oechssler and Riedel (2002) show that neither ESS, CSS nor NIS is sufficient for Lyapunov stability in the weak topology with respect to the replicator dynamics. For doubly symmetric games (including the language game of the present paper) they show that *evolutionary robust* ($\mathcal{ER}$) strategies imply Lyapunov stability in the weak topology with respect to the replicator dynamic.

Definition 4 ($\mathcal{E}$ $\mathcal{R}$, Oechssler and Riedel (2002)) *A (mixed) strategy α is evolutionary robust if $\Lambda(\alpha, \beta) < \Lambda(\beta, \beta)$ for all $\beta \neq \alpha$ that are at least ϵ-close to α in the weak topology.*

Cressman, Hofbauer, and Riedel (2006) give a useful criterion for instability in the replicator equation that can easily be checked for in the present setting. Example 5 demonstrates that a Voronoi language does not need to be Lyapunov stable by checking for this criterion.

Lemma 8 *Evolutionarily robust languages are strict local optima.*

While $\mathcal{ER}$ is a sufficient condition for stability in the weak topology, it is often too strict. As for ESS in the finite case, $\mathcal{ER}$ do not need to exist (see our Example 4 below). We thus look at dynamically stable equilibria next. As the symmetrized sender-receiver game is a game of common interest, the payoff function serves as a Lyapunov function. We thus have

Theorem 3 *Locally optimal languages are Lyapunov stable with respect to replicator (more generally, payoff–monotone) and Brown–von Neumann–Nash dynamics.*

Example 4 below illustrates that an $\mathcal{ER}$ does not need to exist.

Example 4 (($\mathcal{ER}$ do not need to exist)) *For instance, reconsider example 2 (where we have two words and the unit interval as the type space T, plus a uniform probability F distribution over T and a quadratic loss function). The efficient language here is (w^*, i^*) (disregarding null sets and symmetries up to permutation of words) where $w^*(t) = w_1$ if $t \in [0, 1/2]$ and $w(t) = w_2$*

else, and where $i_1^* = 1/4$ *and* $i_2^* = 3/4$. *Now consider a mutant language* $(\hat{w}, \hat{i})$ *with* $\hat{w}^{-1}(w_1) = [0, 1/2+\epsilon]$, $\hat{w}^{-1}(w_1) = (1/2+\epsilon, 1]$, $\hat{i}(w_1) = 1/4+\epsilon$, *and* $\hat{i}(w_2) = 3/4+\epsilon$ *(for some* $\epsilon \in (0, 1/4)$*). We have*

$$\begin{aligned}
L(w,i) &= \int_0^{1/2} (1/4-x)^2 dx + \int_{1/2}^1 (3/4-x)^2 dx \\
&= 1/48 \\
L(w,\hat{i}) &= \int_0^{1/2} (1/4+\epsilon-x)^2 dx + \int_{1/2}^1 (3/4+\epsilon-x)^2 dx \\
&= 1/48+\epsilon^2 \\
L(\hat{w},i) &= \int_0^{1/2+\epsilon} (1/4-x)^2 dx + \int_{1/2+\epsilon}^1 (3/4-x)^2 dx \\
&= 1/48+\epsilon^2/2 \\
L(\hat{w},\hat{i}) &= \int_0^{1/2+\epsilon} (1/4+\epsilon-x)^2 dx + \int_{1/2+\epsilon}^1 (3/4+\epsilon-x)^2 dx \\
&= 1/48+\epsilon^2/2 \\
\Lambda((w,i),(\hat{w},\hat{i})) &= 1/2(L(w,\hat{i}) + L(\hat{w},i)) \\
&= 1/48+3\epsilon^2/4 \\
\Lambda((\hat{w},\hat{i}),(\hat{w},\hat{i})) &= L(\hat{w},\hat{i}) \\
&= 1/48+\epsilon^2/2 < \Lambda((w,i),(\hat{w},\hat{i}))
\end{aligned}$$

As (w, i) *is efficient, a homogenous population of* (w, i)*-players cannot be invaded by a small fraction of mutants of any sort. However, in the weak topology a homogenous population* $(\hat{w}, \hat{i})$ *is also within the* ϵ*-environment of* (w, i). *The calculation above shows that a homogenous population of* $(\hat{w}, \hat{i})$*-players cannot be invaded by a small fraction of* (w, i)*-players either. Considering only these two pure strategies, we are dealing with a* 2×2 *game with the utility matrix*

	(w,i)	$(\hat{w},\hat{i})$
(w,i)	$-\frac{1}{48}$	$-\frac{1}{48}-\frac{3}{4}\epsilon^2$
$(\hat{w},\hat{i})$	$-\frac{1}{48}-\frac{3}{4}\epsilon^2$	$-\frac{1}{48}-\frac{1}{2}\epsilon^2$

In this reduced game both (w, i) *and* $(\hat{w}, \hat{i})$ *are strict equilibria and thus evolutionarily stable. According to a result of Eshel and Sansone (2003), we see that the efficient language is not asymptotically stable, although it is stable.*

As the above example indicates, $\mathcal{E}$volutionary $\mathcal{R}$obustness might be too strong a condition for some type spaces. From Lemma 1 we know that efficient languages exist and from Theorem 1 we know that those languages are Voronoi Languages. As Voronoi Languages are F-a.s. strict Nash equilibria they might seem as candidates for stable states. We prove this wrong by the next example.

Example 5 (A Voronoi language can be unstable)
This example demonstrates that a Voronoi language with full vocabulary does not need to be Lyapunov stable. This might seem surprising as each such language is a strict Nash equilibrium which implies evolutionary stability. As was pointed out earlier, evolutionary stability does not imply stability in games with a continuum of strategies. Consider again example 3 in which the unit square is equipped with the uniform distribution and the loss function is $l(d) = d^2$. *There are two Voronoi languages and therefore two strict Nash equilibria: the horizontal and the diagonal Voronoi tesselation of figure 2. As example 3 points out, the horizontal language is efficient.*

Now we show that the diagonal language is unstable. To specify a mutant strategy we parameterize the equilibrium strategy by the nonnegative small real number a. *The sender mutant strategy is*

$$w_a(t) = \begin{cases} w_1 & \textit{if } t_2 \geq a + (1-2a)\cdot t_1 \\ w_2 & \textit{if } t_2 < a + (1-2a)\cdot t_1 \end{cases}$$

Defining deviating interpretations as functions of a, *we consider best interpretations given* w_a*: Bayesian estimators* $i_a(\hat{w}) = E[t|w_a^{-1}(\hat{w})]$.

$$\begin{aligned} i_a(w_1) &= \left(\frac{1}{3}(1+a), \frac{1}{3}(2+a-a^2)\right) = i(w_1) + \frac{a}{3}\,(1, a(1-a)) \\ i_a(w_2) &= \left(\frac{1}{3}(2-a), \frac{1}{3}(1-a+a^2)\right) = i(w_2) - \frac{a}{3}\,(1, a(1-a)) \end{aligned}$$

We depict this mutant strategy in figure 4. Of course, such a parametrization does not capture all possible deviations. Still, it defines a subset of deviations that can invade a population of agents with the equilibrium strategy in the sense of Apaloo (1997). Further, this parametrization allows us to directly apply Cressman, Hofbauer, and Riedel (2006).

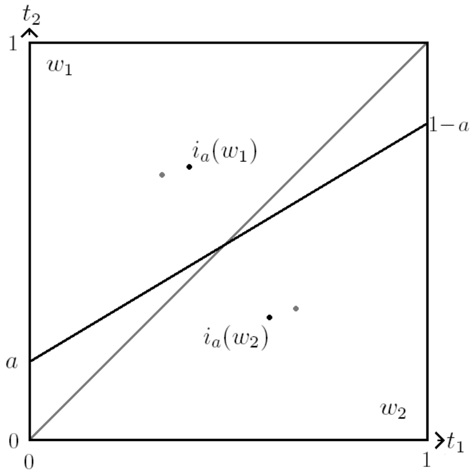


Figure 4: A mutant language

The expected loss of an agent that uses a-deviation (w_a, i_a) and meets an agent that uses b-deviation (w_b, i_b) is then

$$\begin{aligned}
\Lambda(a,b) &= \sum_{\hat{w}\in\{w_1,w_2\}} \frac{1}{2}E\left[||t-i_a(\hat{w})||^2|t\in w_b^{-1}(\hat{w})\right]+\\
&\quad \frac{1}{2}E\left[||t-i_b(\hat{w})||^2|t\in w_a^{-1}(\hat{w})\right]\\
&= \Lambda(0,0)-\\
&\quad \frac{1}{18}\left(2a(1-a)b(1-b)-2\left(b-a\right)^2-\left(b(1-b)-a(1-a)\right)^2\right)
\end{aligned}$$

with gradient

$$\nabla\Lambda(a,b) = -\frac{1}{18}\left[\begin{array}{c} 4(b-a)+2(2b(1-b)-a(1-a))(1-2a) \\ 4(a-b)+2(2a(1-a)-b(1-b))(1-2b) \end{array}\right]$$

and second derivatives

$$\begin{aligned}
\frac{\partial^2\Lambda(a,b)}{(\partial a)^2} &= \frac{1}{9}\left((1-2a)^2+2(1-2b(1-b)-a(1-a))\right)\Big|_0 = \frac{1}{3}\\
\frac{\partial^2\Lambda(a,b)}{\partial a\partial b} &= -\frac{1}{9}\left(2+2(1-2a)(1-2b)\right)\Big|_0 = -\frac{4}{9}\\
\frac{\partial^2\Lambda(a,b)}{(\partial b)^2} &= \frac{1}{9}\left(2+(1-2b)^2+2(2a(1-a)-b(1-b))\right)\Big|_0 = \frac{1}{3}
\end{aligned}$$

At equilibrium $(a, b) = 0$ *the gradient is zero and* $\frac{\partial^2 \Lambda(a,b)}{(\partial a)^2}$ *is positive, which is the analytical implication of that the diagonal language is a strict Nash equilibrium.*

According to Eshel (1983) Theorem 1, a necessary condition for the diagonal language to be continuously stable is that $\frac{\partial^2 \Lambda(a,b)}{(\partial a)^2} + \frac{\partial^2 \Lambda(a,b)}{\partial a \partial b} \geq 0$ *at* $(a, b) = 0$*. As* $\frac{1}{3} - \frac{4}{9} < 0$*, the diagonal language is not CSS.*

Applying Theorem 4 of Cressman, Hofbauer, and Riedel (2006) to our setting, $\frac{\partial^2 \Lambda(a,b)}{(\partial a)^2} + \frac{\partial^2 \Lambda(a,b)}{\partial a \partial b} < 0$ *at the diagonal equilibrium* $(a, b) = 0$ *implies that the state in which each agent of the population chooses* (w_0, i_0) *is unstable with respect to the replicator equation restricted to normal distributions.*

One can check that $\Lambda(0, a) > \Lambda(a, a)$*, hence the diagonal language is neither NIS (Apaloo (1997)) nor* $\mathcal{ER}$ *(Oechssler and Riedel (2002)).*

Note that $\Lambda(\cdot, \cdot)$ *denotes losses, hence we need to consider reverse inequalities of the cited literature.*

In the example above, we showed instability by parameterizing appropriate mutation strategies that can invade the diagonal language. The next example in turn shows (Lyapunov) stability of languages on a rectangle type space by deriving the property of local optimality. We can of course show by the same means that the diagonal language is not a local optimum and will do so at the end of the next example.

Example 6 (A stable Voronoi language can be inefficient)
This example demonstrates that not every stable equilibrium is optimal. Consider a rectangle A where the sides have length a and b respectively. We consider a type space where each point in the interior of A has a uniform probability density 1, and all other points have probability density 0. Also, we assume a quadratic loss function as in the other examples. There are two words, w_1 *and* w_2*.*

Both bipartitions of the rectangle that split A into two identical rectangles along a boundary that runs parallel to one pair of sides (together with the centers of the partition cells as prototypes) represent stable languages (see Figure 5). However, only the partition that is parallel to the short sides is optimal.

To prove that the other partition is stable but not optimal, we first show that both partitions are local minima. According to Theorem 3, they are thus both stable. Finally, we show that the loss of the second partition (right hand side of Figure 5) is strictly larger than the loss of the first partition.

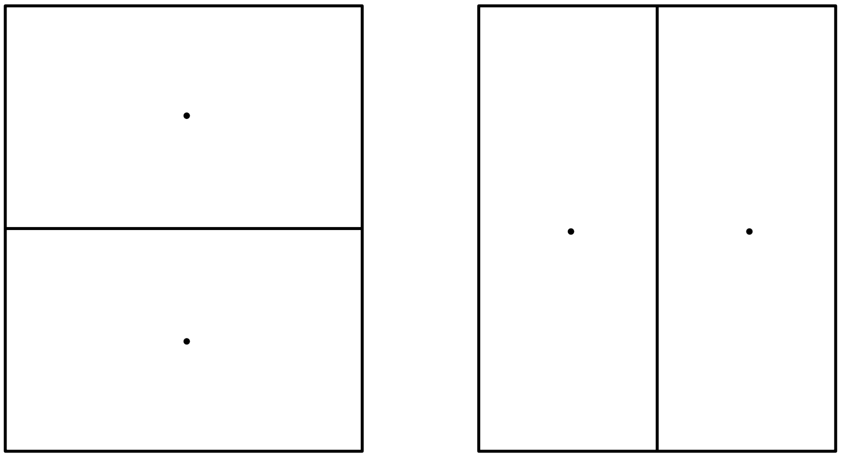


Figure 5: Stable partitions of a rectangle

We start with the proof that both partitions are local minima. To do so, we embed A into a Cartesian coordinate system, with the four corners located at $(0,0)$, $(a,0)$, $(0,b)$, and (a,b). This is depicted in Figure 6. For the time being, we make no assumptions whether $a < b$ or $a > b$. Consider the language (w^*, i^*), where $w^*((x,y)) = w_1$ iff $x \in (0,a)$ and $y \in (0, \frac{1}{2}b)$, $w^*((x,y)) = w_2$ iff $x \in (0,a)$ and $y \in [\frac{1}{2}b, b)$, $i_1^* = (\frac{1}{2}a, \frac{1}{4}b)$ and $i_2^* = (\frac{1}{2}a, \frac{3}{4}b)$. We will prove now that this language is a local minimum.

Suppose it is not a local minimum. Then there is a sequence of languages (w_k, i_k) that converges toward (w^*, i^*) such that for some k^*, for all $k' > k^* : L(w_{k'}, i_{k'}) \leq L(w^*, i^*)$. The best response $BR(i_{k'})$ is the sender strategy that is the best response to the receiver strategy $i_{k'}$.[14] *It is the Voronoi partition that is induced by $i_{k'}$. Because the best response function is continuous, the sequence $(BR(i_{k'}), i_{k'})$ also converges towards (w^*, i^*). If $L(w_{k'}, i_{k'}) \leq L(w^*, i^*)$, it also holds that $L(BR(i_{k'}), i_{k'}) \leq L(w^*, i^*)$. To show that this is impossible, it is sufficient to show that $L(w^*, i^*) > L(BR(i), i)$ for all $i \neq i^*$ in some ϵ-environment of i^*.*

Let us assume that $i_1 = (x_1, y_1)$ and $i_2 = (x_2, y_2)$ (as indicated in Figure 6). The line that separates $w^{-1}(w_1)$ from $w^{-1}(w_2)$ (with $w = BR(i)$ is given

[14] The best response to a receiver strategy $i_{k'}$ is unique—up to the images of null sets—if $i_{k',1} \neq i_{k',2}$. Since we are considering strategies in the environment of i^*, we can safely assume that this is the case.

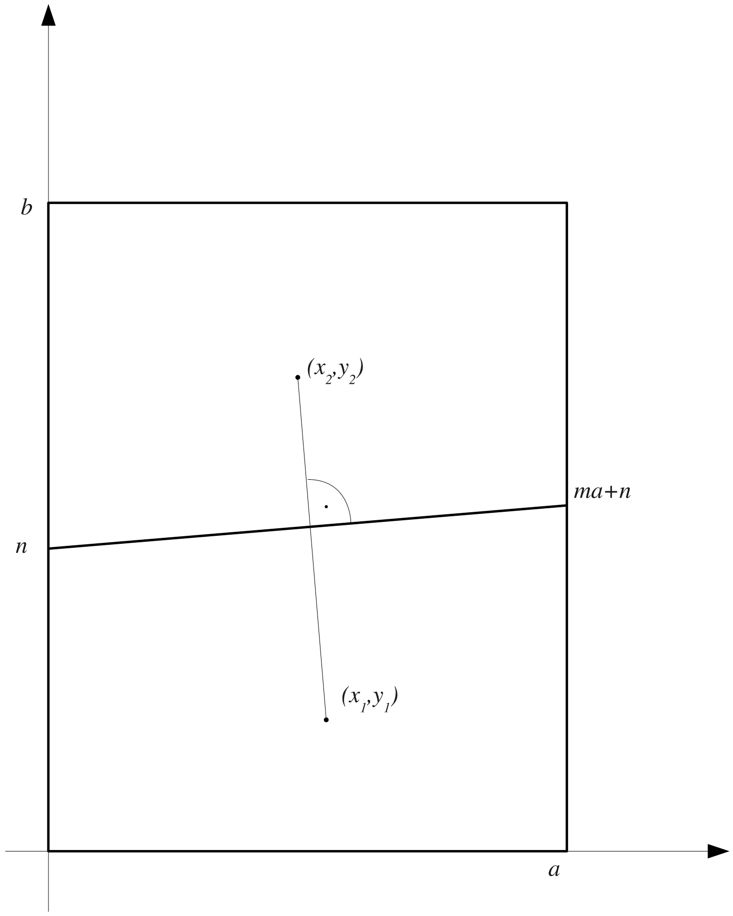


Figure 6:

by the function $g(x)$*, where*

$$\begin{aligned} g(x) &= mx + n, \\ m &= \frac{x_1 - x_2}{y_2 - y_1}, \\ n &= \frac{x_2^2 + y_2^2 - x_1^2 - y_1^2}{2(y_2 - y_1)}. \end{aligned}$$

(We assume here that $y_1 \neq y_2$*, which obviously holds in the environment of* i^**. We also assume that* $g(x)$ *intersects with both the y-axis and the line* $y = a$ *in the interval* $[0, b]$*, which is also true in the environment of* i^**.) So*

the loss $L(BR(i), i)$ is given by

$$\begin{aligned} L(BR(i), i) &= \int_0^a \int_0^{g(x)} (x - x_1)^2 + (y - y_1)^2 dy dx \\ &\quad + \int_0^a \int_{g(x)}^b (x - x_2)^2 + (y - y_2)^2 dy dx. \end{aligned}$$

The gradient of this function in the four-dimensional parameter space defined by x_1, y_1, x_2 and y_2 is

$$\begin{aligned} \left.\frac{\partial L(BR(i), i)}{\partial x_1}\right|_{i=i^*} &= 0 \\ \left.\frac{\partial L(BR(i), i)}{\partial y_1}\right|_{i=i^*} &= 0 \\ \left.\frac{\partial L(BR(i), i)}{\partial x_2}\right|_{i=i^*} &= 0 \\ \left.\frac{\partial L(BR(i), i)}{\partial y_2}\right|_{i=i^*} &= 0, \end{aligned}$$

so i^* is a critical point.

The Hessian of the loss function at i^* is

$$\begin{pmatrix} ab - \frac{a^3}{3b} & \frac{a^3}{3b} & 0 & 0 \\ \frac{a^3}{3b} & ab - \frac{a^3}{3b} & 0 & 0 \\ 0 & 0 & \frac{3}{4}ab & -\frac{1}{4}ab \\ 0 & 0 & -\frac{1}{4}ab & \frac{3}{4}ab \end{pmatrix}.$$

The eigenvalues of this matrix are $\frac{3ab^2-2a^3}{3b}$, ab, and $\frac{ab}{2}$. So if $a < b\sqrt{\frac{3}{2}}$, the matrix is positive definite, and i^* is in fact a local minimum of the loss function. Therefore i^* is stable. We express the gradient and the Hessian as functions of $i(w_1)$ and $i(w_2)$ explicitly in the appendix.

Now assume $b\sqrt{\frac{2}{3}} < a < b$. It holds that

$$L(w^*, i^*) = \frac{1}{48}(ab^3 + 4a^3b).$$

Let (w^{**}, i^{**}) be the equilibrium with $i_1^{**} = (\frac{1}{4}a, \frac{1}{2}b)$ and $i_2^{**} = (\frac{1}{4}a, \frac{1}{2}b)$, and w^{**} the Voronoi tesselation induced by i^{**}. By the argument given above, it *is also stable. (You only have to exchange a and b, and x and y in the proof above.) Here we have*

$$L(w^{**}, i^{**}) \;\; = \;\; \frac{1}{48}(a^3b + 4ab^3).$$

Some elementary calculations reveal that $L(w^*, i^*) < L(w^{**}, i^{**})$ iff $a < b$. So if $b\sqrt{\frac{2}{3}} < a < b$, $L(w^{**}, i^{**})$ is stable, but it is not optimal.

Let us now briefly reconsider the diagonal language from the previous example. The diagonal language solves the first order conditions $L'(BR(i), i) = 0$ only if $a = b$, in other words if the rectangle is a square. Plugging in the diagonal language into the Hessian of the loss function reveals that the diagonal language is a saddle point and not a local optimum. Again, see the appendix for explicit details.

6 An Algorithm for Computing Voronoi Languages and Further Examples

In this section we examine languages with more than two words. As the computational problem of solving for three Voronoi tiles is demanding the problem becomes ambitious for more than three words. Even more challenging is to analyze stability properties. In the unit square example, a language with three words can be parametrized by a six-dimensional vector (two 'coordinates' for each of the three interpretations). For stability analysis, one needs to calculate the 6×6 Hessian matrix of the loss function. At least for more complex languages we expect to loose tractability when following a strictly analytical approach. For this reason we provide a section that relies on simulations. Although not stringent from a mathematical viewpoint, such simulations can well indicate whether a particular Voronoi tessellation is stable or not. Further, one can extend the algorithm described here easily to settings with a finite–dimensional state space T or to a setting with arbitrary distribution functions.

6.1 The Algorithm

We describe the algorithm step by step. The source code can be found in the appendix.

- *Initialization $t = 0$:* $\quad i_1(0), i_2(0), \ldots, i_N(0)$

  To start the algorithm, the interpretations receive initial values. These can be chosen sophistically as a particular Voronoi tessellation to test for its robustness in the presence of randomness. Alternatively, they can be randomly assigned to check for path dependence.

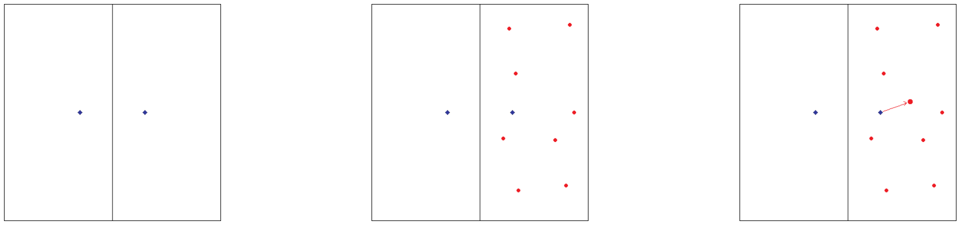
Hereafter, the algorithm finitely often iterates the following two steps:

- *Random Types*

  Each iteration begins with randomly drawing finitely many types from T. Each sensation is assigned to its closest interpretation.

- *Tile Adaption*

  The new value of the interpretation that represents a tile is the arithmetic mean of the types that are contained in that tile.

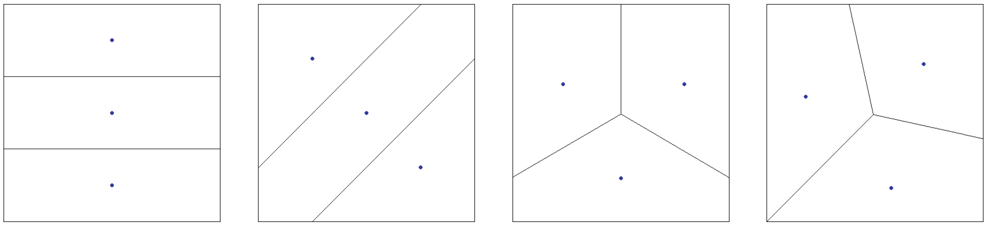


Initialization *Random Types* *Tile Adaption*

This surprisingly simple algorithm robustly selects some particular languages from a variety of Voronoi languages. On the other hand it is easy to show that some candidate languages render unstable in the presence of small deviations. We give some examples:

6.2 Three Words

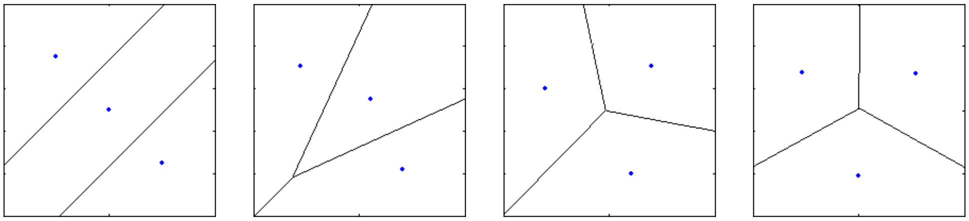
If the language comprises three words, up to symmetry there are two types of Voronoi tessellations which each have a 'horizontal' and a 'diagonal' version:



Ia Ib IIa IIb

The algorithm selects language ‘IIa’ which also has the property of minimizing the expected loss. This has been tested for arbitrary initial conditions. The appendix derives the tessellations analytically, nevertheless we need to rely on the simulations for the finding of stability.

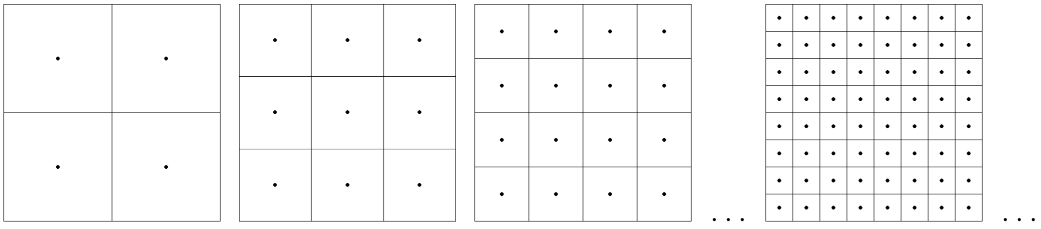
The figures below show four snapshots of a simulation that starts at the equilibrium ‘Ib’. The process quickly leaves ‘Ib’. The lines separating the three categories of the equilibrium have merged at time $t = 50$ and from $t = 100$ on the process freezes for some time in equilibrium ‘IIb’. But as the initial state, this Voronoi Language does not render stable either. In the long run (from $t = 1000$), the process reaches equilibrium ‘IIa’ which is persistent.



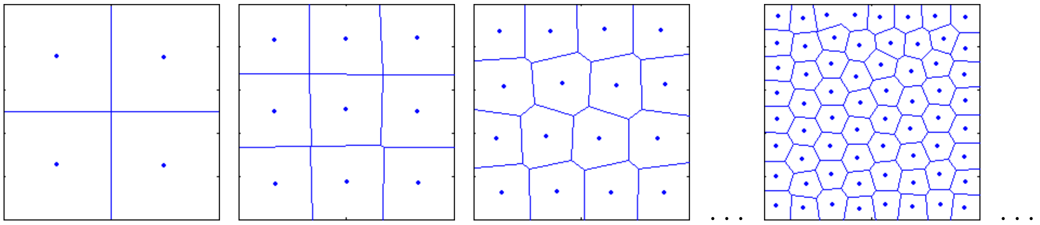
$t = 1$ $t = 50$ $t = 100$ $t = 1000$

6.3 Triangles, Squares and Hexagons

As indicated above, characterizing the set of Voronoi tessellations becomes more complex a problem, the more words the language has at disposal. Still, some tessellations are straightforward to describe. For any $n \in \mathbb{N}$, there is a Voronoi language with n^2 cells, as is illustrated below.



For small n, these languages are stable while for large n they are not; indeed evolution leads to a hexagonal structure. We depict the tessellations after 1000 iterations.



All examples that have been presented up to now have in common that the borders of the type space have an impact on equilibrium tesselations. One way to prevent this is considering type distributions with mass close to zero near the boundary. We follow this path in the next subsection. In this subsection we circumvent boundary effects by considering unbounded or boundaryless type spaces. Imagine a square whose opposite edges are stuck together. If an interpretation that is located at the south east corner moves further in direction south east, it will appear in the north western corner of the square.[15] We conjecture that an efficient tesselation for this square without boundaries does consist of regular polygons, that is any two adjacent sides do have the same interior angle and have the same length.[16] For any

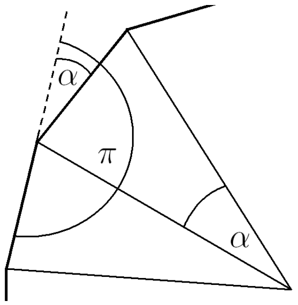


Figure 7: The interior angle of a rectangular polygon equals $\pi - \alpha$, where α equals 2π divided by the number of vertices.

[15] More precisely one should consider a torus (which is compact) and its projection onto the plane. To keep things transparent we leave these concerns aside.

[16] We are aware of 'stripe languages' for which the interpretations lie on equidistant points on a straight line, which do not belong to the polygon languages. These languages can easily be shown to be unstable.

convex polygon with v vertices, the interior angle of a regular polygon equals $\pi - \frac{2\pi}{v}$, as can be seen from figure 7 with $\alpha = \frac{2\pi}{v}$. Within a tesselation, the number of edges that a vertex can have is then $\frac{2\pi}{\pi - \frac{2\pi}{v}} = \frac{2v}{v-2}$ which is an integer only for $v = 3, 4$ and 6. We conclude that there are only three regular polygons that can cover the boundary-less square entirely: triangles, squares and hexagons. If we compare the expected losses of a triangle-, a square- and a hexagon-language with n words, the hexagon-language has the lowest expected loss.

For uniformly distributed types, each of the n regular cells has mass $\frac{1}{n}$. The length of an edge of a regular triangle is then $\frac{2}{\sqrt{n\sqrt{3}}}$, see figure 8. The

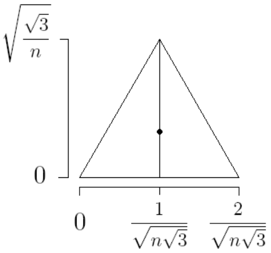

$\sqrt{\frac{\sqrt{3}}{n}}$
0
0
$\frac{1}{\sqrt{n\sqrt{3}}}$
$\frac{2}{\sqrt{n\sqrt{3}}}$

Figure 8: A regular triangle of mass $\frac{1}{n}$

expected loss of such a triangular language with n words is

$$\Lambda(\triangle) = 2n \int_0^{\frac{1}{\sqrt{\sqrt{3}n}}} \int_0^{\sqrt{3}t_1} \left(t_1 - \frac{1}{\sqrt{n\sqrt{3}}}\right)^2 + \left(t_2 - \frac{1}{3}\sqrt{\frac{\sqrt{3}}{n}}\right)^2 dt_2 dt_1 = \frac{1}{3\sqrt{3}n}$$

For the square-language with n words, each having categories of area $\frac{1}{n}$ (and side-length $\frac{1}{\sqrt{n}}$) we have

$$\Lambda(\square) = n \cdot 4 \int_0^{\frac{1}{2\sqrt{n}}} \int_0^{\frac{1}{2\sqrt{n}}} \left(t_1 - \frac{1}{2\sqrt{n}}\right)^2 + \left(t_2 - \frac{1}{2\sqrt{n}}\right)^2 dt_2 dt_1 = \frac{1}{6}\frac{1}{n}$$

We depict a regular hexagon with mass $\frac{1}{n}$ in figure 9 below. The expected loss of a hexagon-language with n words is

$$\Lambda(\text{⬡}) = n \cdot 12 \int_0^{\frac{1}{\sqrt{6n\sqrt{3}}}} \int_0^{t_1\sqrt{3}} \left(t_1 - \frac{1}{\sqrt{6n\sqrt{3}}}\right)^2 + \left(t_2 - \frac{1}{\sqrt{2n\sqrt{3}}}\right)^2 dt_2 dt_1 = \frac{5}{18}\frac{1}{n\sqrt{3}}$$

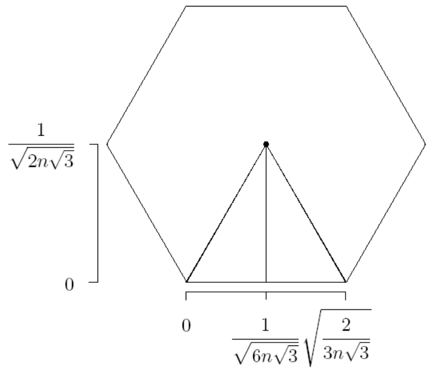

$\frac{1}{\sqrt{2n\sqrt{3}}}$
0
0
$\frac{1}{\sqrt{6n\sqrt{3}}}$
$\sqrt{\frac{2}{3n\sqrt{3}}}$

Figure 9: A regular hexagon of mass $\frac{1}{n}$

Summing up, a language with a large number of words has lower expected loss, if the shape of the categories is hexagonal: $\Lambda(\triangle) > \Lambda(\square) > \Lambda(\text{⬡})$.

Still the efficient language (and by same arguments any Voronoi language) fails $\mathcal{E}$volutionary $\mathcal{R}$obustness. Consider a homogenous population of agents that use the hexagonal language and small group of invading mutants who use the hexagonal language shifted slightly as depicted in the figures below. Denote the original hexagonal language by 0 and the mutants by a. Clearly

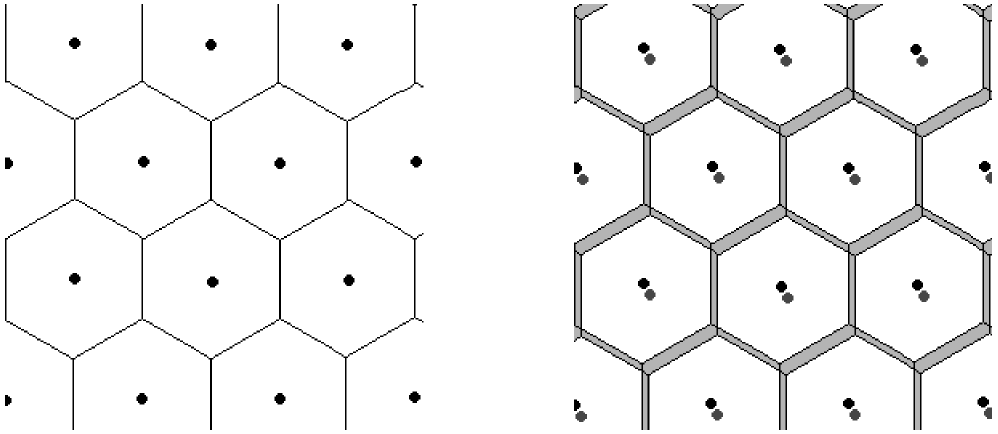


Figure 10: The hexagonal language invaded by shifted mutants

we have $\Lambda(0, 0) = \Lambda(a, a)$. However, when an agent with the original language meets a mutant, there will be misunderstanding for types in the shaded area.

Hence $\Lambda(0,a) > \Lambda(0,0)$ and hereby also $\Lambda(0,a) > \Lambda(a,a)$. Therefore, the requirement for $\mathcal{E}$volutionary $\mathcal{R}$obustness is not met.

6.4 Other Type Distributions

Let us have a brief view at non-uniform type distributions. Figure 11 shows normally distributed types. Having the variance small enough one can hereby simulate a tesselation without 'border effects' as the border of the type space receives mass close to zero. Hence the resulting tesselation is invariant to the shape of the type space. Examples of such a distribution would be parameters that realize within natural boundaries such as weather conditions (temperature, humidity) or traffic conditions (the speed of an approaching vehicle or the crowdedness of a particular highway) or economic parameters (like prices, profits or probabilities). One can observe that the tesselation approximates the hexagonal structure around the mean of the normally distributed types.

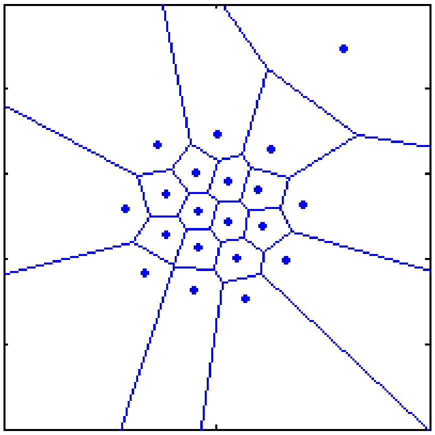


Figure 11: Tesselation of normally distributed types

Figure 12 represents a setting in which types realize more often close to one boundary of the type space than the other. For example the color specifier *red* (1.110 million Google hits) seems to have a drastically higher intensity of usage than the word *yellow* (455 million Google hits). Note that regions with lower mass have large categories and that smaller categories are found where mass is concentrated.

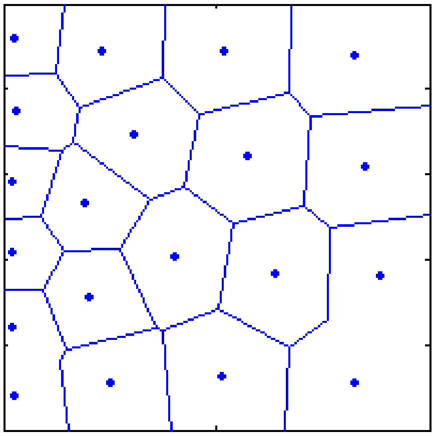


Figure 12: Tesselation of asymmetrically distributed types

6.5 Relation to k-means clustering

The algorithm described above can be seen as a stochastic generalization of the k-means clustering algorithm that is widely used in multivariate statistical data analysis and machine learning (sometimes under the heading of *vector quantization*; see for instance chapter 9.1 in Bishop (2006)). In these applications, we have finitely many observations that are unevenly distributed in an L-dimensional Euclidean space. The goal is to find a partition of the observations into k clusters (for some natural number k that is small in comparison to the number of observations) that minimizes the within-cluster loss (squared distance) and maximizes the between-cluster loss. The standard algorithm to find an optimal clustering of this kind is to start with an arbitrary k-tuple of prototypes, calculate the corresponding Voronoi tessellation, and to update each prototype to the arithmetic mean of the observations within its Voronoi tile. This process is repeated until a fixed point is reached. In the language of game theory, this amounts to an iterated best response computation for a discrete probability distribution of the type space. Our algorithm generalizes this idea to the continuous case.

7 Conclusion

We analyze common interest signaling games when the type space is much more complex than the signal set. Efficient signaling systems (”languages”) use Voronoi tesselations of the type space to transmit information. These Voronoi tesselations must also satisfy a best estimator property: the prototypes that generate the Voronoi tesselations form a best estimator in the Bayesian sense. We have seen that these "Voronoi languages" are exactly the strict Nash equilibria of the signaling game. While not all Voronoi languages are dynamically stable under replicator or similar evolutionary dynamics, efficient languages are. Nevertheless, evolution can also converge to inefficient Voronoi languages.

A Proofs

A.1 Existence of Efficient Languages (Lemma 1)

We can identify strategies $w : T \to W$ for the sender with the corresponding partition $(C_j)_{j=1,\dots,N}$ given by

$$C_j = \{t \in T | w(t) = w_j\} \ .$$

Let $(i_j)_{j=1,\dots,N}$ be a pure strategy for the receiver. Given that strategy, a type t optimally selects a word that leads to an interpretation that is as close as possible to t, i.e. $w(t) \in \operatorname{argmin}_{j=1,\dots,N} \|t - i_j\|$. Note that in general, the interpretations i_j need not be pairwise distinct. In this case, we choose always the index with the smallest subscript, so we set

$$C_k^i = \left\{ t \in T | k \text{ is the smallest number in } \operatorname{argmin}_{j=1,\dots,N} \|t - i_j\| \right\} \ .$$

We have thus reduced our optimization to a minimization problem over the compact set T^N, namely

$$\min_{(i_j)_{j=1,\dots,N} \in T^N} \int_T \sum_{k=1}^N l\left(\|t - i_k\|\right) 1_{C_k^i}(t) F(dt) \, .$$

For the existence of an efficient language, it is thus sufficient to prove the continuity of the integral

$$\int_T \sum_{k=1}^N l\left(\|t - i_k\|\right) 1_{C_k^i}(t) F(dt)$$

in (i_k). By Lebesgue's theorem of dominated convergence, it is enough to show that the integrand $\sum_{k=1}^N l\left(\|t-i_k\|\right) 1_{C_k^i}(t)$ is F–almost everywhere continuous. We can ignore the boundaries of the sets C_k^i because these boundaries are intersections of hyperplanes with the set T and therefore Lebesgue, hence F–nullsets. Take a type $t \in T$ in the interior of some C_m^i for some $1 \leq m \leq N$. Being in the interior of C_m^i, i_m is the unique interpretation with minimal distance to t. Take a sequence $\left((i_j^n)_{j=1,\ldots,N}\right)_{n\in\mathbb{N}}$ with $i_j^n \to i_j$ as $n \to \infty$ for all $j = 1, \ldots, N$. For n sufficiently large, i_m^n is the unique interpretation among (i_j^n) with minimal distance to t and $i_k^n \in C_k^i$. Therefore, the continuity of l entails

$$\sum_{k=1}^N l\left(\|t-i_k\|\right) 1_{C_k^i}(t) = l\left(\|t-i_m\|\right) 1_{C_m^i}(t) = \lim_{n\to\infty} l\left(\|t-i_m^n\|\right) 1_{C_m^i}(t)\,.$$

Thus, the integrand is F-a.e. continuous.

A.2 Mixed Strategies are Never Efficient (Lemma 2)

Fix any $t \in T$. Randomized strategies (ω, μ) lead to a probability distribution $\gamma_t(di) = \sum_{k=1}^N \mu_k(di)\omega_k(t)$ over T. Suppose that this measure is not a Dirac measure.

Now denote by $\bar{\gamma} = \sum_{k=1}^N \int_T i\ \mu_k(di)\omega_k(t)$ the average outcome of communication in T when (ω, μ) is played. The function $i \mapsto l\left(\|t-i\|\right)$ is strictly convex; by Jensen's inequality,

$$\sum_{k=1}^N \int_T l\left(\|t-i_k\|\right) \mu_k(di_k) \leq l\left(\|t-\bar{\gamma}\|\right)\,,$$

and the inequality is strict when γ is not a Dirac measure. This shows that mixing is never efficient.

A.3 Structure of Efficient Languages (Lemma 3, Lemma 4, Lemma 5) and Theorem 1

Let (w, i) be an efficient language. From our analysis in Section A.1, we know that we can identify w without loss of generality with the partition

$$C_k = \left\{t \in T | k \text{ is the smallest number in } \operatorname{argmin}_{j=1,\ldots,N} \|t-i_j\|\right\}\,.$$

Note that these sets C_k are either intersections of convex polyhedra with the type set T or empty, if some word is not used. Suppose that the word w_N is not used, i.e. $C_N = \emptyset$. The idea of the proof is to take a word that is used for a big set of types and to split that set into two smaller sets and to use two words instead of one. This allows to decrease the average loss.

By definition, word w_1 is used with positive probability , i.e. the convex set C_1 has positive mass with respect to F. As F is atomless, we can find two disjoint, convex, nonnull sets A_1, A_N with $A_1 \cup A_N = A$. Now let $j_1 \in T$ be a minimizer[17] of

$$\int_{A_1} l\left(\|t-j\|\right) F(dt),$$

and similarly, $j_N \in T$ be a minimizer of

$$\int_{A_N} l\left(\|t-j\|\right) F(dt).$$

By strict convexity of l, the minimizers are uniquely determined. Moreover, we have $j_1 \neq j_N$ because the minimizer lies in the interior of A_1 resp. A_N.

Set $j_k = i_k$ for $k = 2, \ldots, N-1$. Moreover, set $v(t) = w_1$ for $t \in A_1$ and $v(t) = w_N$ for $t \in A_N$, and $v(t) = w(t)$ else. We claim that (v, j) is a better language than (w, i):

$$\begin{aligned} L(v,j) - L(w,i) = & \int_{A_1} \left(l\left(\|t-j_1\|\right) - l\left(\|t-i_1\|\right)\right) \\ & + \int_{A_N} \left(l\left(\|t-j_N\|\right) - l\left(\|t-i_1\|\right)\right) > 0 \end{aligned}$$

where the last inequality comes from the fact that j_1 and j_N minimize the loss over the sets A_1 and A_N and either $j_1 \neq i_1$ or $j_N \neq i_1$.

It remains to be shown that all interpretations (i_k) are pairwise distinct. Given that the signaling system is induced by a partition (C_k) of convex sets with positive measure, the optimal interpretation for word w_k is the "prototype" i_k that minimizes

$$\int_{C_k} l\left(\|t-j\|\right) F(dt)$$

[17] The minimum exists because T is compact and the expression is continuous in j, see the proof of Lemma 1.

for $j \in T$. As C_k is convex and F atomless, the minimizer lies in the interior of the set C_k. In particular, all interpretations (i_k) are pairwise distinct for an efficient language. Moreover, we see that the receiver uses a best estimator in the sense of Definition 2.

A.4 Evolution (Proof of Lemma 6, Theorem 3, Lemma 8)

The proof that average loss is decreasing along payoff–monotone dynamics and BNN dynamics follows well–known lines, see Heifetz, Shannon, and Spiegel (2007) and Hofbauer, Oechssler, and Riedel (2009). The loss function Λ is continuous with respect to the weak topology if the direct loss function for pure strategies $L(w,i)$ is continuous (in the usual norm on T^N and Σ endowed with the supremum–norm) and bounded.

The maximal distance on T is bounded because T is compact, and l is continuous, so L remains bounded.

To see continuity, choose a sequence (w^n) of sender strategies that converge uniformly to w and a sequence (i^n) of receiver strategies that converges to $i \in T^N$. Let $\epsilon > 0$ and $\delta > 0$ such that $|l(d) - l(e)| < \epsilon$ for all $0 \le d, e \le \max_{s,t\in T} \|s-t\|$. (Note that l is uniformly continuous on bounded intervals and that the maximum is finite because T is compact.) As sender strategies can assume only finitely many values in W, there exists $N_0 \in \mathbb{N}$ such that $w^n(t) = w(t)$ uniformly in $t \in T$ for $n \ge N_0$. It follows that

$$\|t - i_{w^n(t)}\| = \|t - i_{w(t)}\| \tag{4}$$

uniformly in $t \in T$ for $n \ge N_0$. Now choose $N_1 \ge N_0$ such that for $n \ge N_1$

$$\|i^n_j - i_j\| < \delta$$

for all $j \in \{1, \dots, N\}$. Then

$$\begin{aligned}
|L(w^n, i^n) - L(w,i)| &\le \int_T \left| l\left(\|t - i^n_{w^n(t)}\|\right) - l\left(\|t - i^n_{w(t)}\|\right)\right| F(dt) \\
\text{Eqn.(4)} \quad &= \int_T \left| l\left(\|t - i^n_{w(t)}\|\right) - l\left(\|t - i^n_{w(t)}\|\right)\right| F(dt) \\
\text{(Def. of } \delta) \quad &< \epsilon\,.
\end{aligned}$$

Hence, L is continuous.

Let P^* be evolutionarily robust with invasion barrier $\epsilon > 0$. Then we have for populations $Q \neq P^*$

$$\begin{aligned}\Lambda(P^*,P^*) - \Lambda(Q,Q) &= \Lambda(P^*,P^*) - \Lambda(P^*,Q) + \Lambda(P^*,Q) - \Lambda(Q,Q) \\ &< \Lambda(P^*,P^*) - \Lambda(P^*,Q) \\ &= \Lambda(P^*,P^*) - \Lambda(Q,P^*) \leq 0,\end{aligned}$$

where we use the definition of ER, symmetry of Λ and the fact that (P^*,P^*) is a Nash equilibrium. This shows that P^* is a strict local minimum of $\Lambda(Q,Q)$ and proves Lemma 8.

Let us come to stability questions (Theorem 3). With a Lyapunov function, dynamic stability of local optima follows as usual. We restate this result from Bhatia and Szegő (1970):

Theorem 4 (Bhatia and Szegő (1970) Theorem 2.2) *[Let X be locally compact.] A compact set $M \subset X$ is asymptotically stable if and only if there exists a continuous real-valued function Φ defined on a neighborhood N of M such that*

$\Phi(x) = 0$ *if* $x \in M$ *and* $\Phi(x) > 0$ *if* $x \notin M$*;*
$\Phi(xt) < \Phi(x)$ *for* $x \notin M$, $t > 0$ *and* $x[0,t] \subset N$*.*

For our purposes $X = T^N \times \Sigma$ and $\Phi(x) = \gamma - \Lambda(x)$, where γ is an appropriately chosen constant.

A.5 The line [0,1]

We consider now finite languages on real intervals. For simplicity, we look at uniformly distributed types and quadratic loss.

The game has many symmetries. In particular, for every Voronoi language, there exists an isomorphic language in which the words are permuted arbitrarily. Without loss of generality, we thus look at Voronoi languages that consist of points $0 \leq i_1 < i_2 < \ldots < i_K \leq 1$ for the receiver and corresponding Voronoi cells $[b_0,b_1),[b_1,b_2],\ldots,[b_{K-1},b_K]$ for the sender, with $b_0 = 0$ and $b_K = 1$. $K \leq N$ is the richness of the language.

We claim that we must have

$$b_i = \frac{i}{K},\; i_j = \frac{2j-1}{2K}\,.$$

In other words: *Voronoi languages on* $[0,1]$ *consist of equidistant partitions and their midpoints.* Up to symmetries, there exists only one Voronoi language of a given richness.

PROOF : As the interpretation is the conditional expected types in a cell, we must have $i_j = \frac{b_{j-1}+b_j}{2}, j = 1, \dots, K$. On the other hand, the points $b_0, \dots, b_K$ describe the Voronoi tessellation corresponding to $i_1, \dots, i_K$. Hence, we must have

$$i_1 = \frac{b_1}{2}, i_2 = \frac{b_2 + b_1}{2}, \dots, i_K = \frac{b_{K-1} + 1}{2}\,. \tag{5}$$

The unique solution of this system of linear equations is

$$b_i = \frac{i}{K},\ i_j = \frac{2j-1}{2K}\,.$$

It is straightforward to see that this is a solution. Uniqueness may be unclear. Note that the i_j are uniquely determined by the b_l. Replace i_j by $1/2(b_{j-1} + b_j)$ in Eqn. (5). Rearrange these equations and you get sequentially

$$\begin{aligned} b_2 &= 2b_1 \\ b_3 &= 2b_2 - b_1 = 3b_1 \\ b_4 &= 4b_1 \\ &\vdots \end{aligned}$$

and so on until $b_K = Kb_1 = 1$ and you are done. □

A.6 Rectangle Type Space

Considering a rectangle of width $a > 0$ and height $b > 0$, denote the interpretations as $i(w_1) = (i_1(w_1), i_2(w_1))$ and $i(w_2) = (i_1(w_2), i_2(w_2))$. Assume $i_2(w_2) > i_2(w_1)$. Given a realized type $t = (t_1, t_2)$, the senders best reply to these interpretations is

$$BR(t|i) = \begin{cases} w_1 & \text{if } t_2 \le g(t_1) \\ w_2 & \text{if } t_2 > g(t_1) \end{cases}, \text{ where}$$

$$g(t_1) = \frac{i_2^2(w_1) - i_1^2(w_1) + i_2^2(w_2) - i_1^2(w_2)}{2(i_2(w_2) - i_2(w_1))} - \frac{i_2(w_1) - i_1(w_1)}{i_2(w_2) - i_1(w_2)} t_1$$

is an affine function in t_1.

After solving the integrals, the loss function can be expressed as

$$L(i, w_i) = -(i_2(w_2) - i_2(w_1)) \int_0^a g^2(t_1) dt_1 + \left(\frac{a^2}{3} - i_1(w_2)a + i_1^2(w_2) + \frac{b^2}{3} - i_2(w_2)b + i_2^2(w_2) \right) ab$$

For the reader's convenience we express the first derivatives of the separating function $g(\cdot)$:

$$\begin{aligned} \frac{\partial g(t_1)}{\partial i_1(w_1)} &= \frac{t_1 - i_1(w_1)}{i_2(w_2) - i_2(w_1)} \\ \frac{\partial g(t_1)}{\partial i_1(w_2)} &= \frac{i_1(w_2) - t_1}{i_2(w_2) - i_2(w_1)} \\ \frac{\partial g(t_1)}{\partial i_2(w_1)} &= \frac{g(t_1) - i_2(w_1)}{i_2(w_2) - i_2(w_1)} \\ \frac{\partial g(t_1)}{\partial i_2(w_2)} &= \frac{i_2(w_2) - g(t_1)}{i_2(w_2) - i_2(w_1)} \end{aligned}$$

The first derivatives of the loss function with respect to the interpretations:

$$\begin{aligned} \frac{\partial \Lambda((BR(i), i), (BR(i), i))}{\partial i_1(w_1)} &= \int_0^a g(t_1) 2(i_1(w_1) - t_1) dt_1 \\ \frac{\partial \Lambda((BR(i), i), (BR(i), i))}{\partial i_1(w_2)} &= \int_0^a g(t_1) 2(t_1 - i_1(w_2)) dt_1 + (2i_1(w_2) - a)ab \\ \frac{\partial \Lambda((BR(i), i), (BR(i), i))}{\partial i_2(w_1)} &= \int_0^a (2i_2(w_1) - g(t_1)) g(t_1) dt_1 \\ \frac{\partial \Lambda((BR(i), i), (BR(i), i))}{\partial i_2(w_2)} &= \int_0^a g(t_1)(g(t_1) - 2i_2(w_2)) dt_1 + (2i_2(w_2) - b)ab \end{aligned}$$

The second derivatives of the loss function with respect to the interpretations:

$$\begin{aligned}
\frac{\partial^2\Lambda((BR(i),i),(BR(i),i))}{\partial i_1(w_1)\partial i_1(w_1)} &= 2\int_0^a g(t_1) - \frac{(t_1-i_1(w_1))^2}{i_2(w_2)-i_2(w_1)}dt_1 \\
\frac{\partial^2\Lambda((BR(i),i),(BR(i),i))}{\partial i_1(w_1)\partial i_1(w_2)} &= 2\int_0^a \frac{i_1(w_2)-t_1}{i_2(w_2)-i_2(w_1)}(i_1(w_1)-t_1)dt_1 \\
\frac{\partial^2\Lambda((BR(i),i),(BR(i),i))}{\partial i_1(w_1)\partial i_2(w_1)} &= 2\int_0^a \frac{g(t_1)-i_2(w_1)}{i_2(w_2)-i_2(w_1)}(i_1(w_1)-t_1)dt_1 \\
\frac{\partial^2\Lambda((BR(i),i),(BR(i),i))}{\partial i_1(w_1)\partial i_2(w_2)} &= 2\int_0^a \frac{i_2(w_2)-g(t_1)}{i_2(w_2)-i_2(w_1)}(i_1(w_1)-t_1)dt_1
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2\Lambda((BR(i),i),(BR(i),i))}{\partial i_1(w_2)\partial i_1(w_2)} &= 2\int_0^a b - \frac{(i_1(w_2)-t_1)^2}{i_2(w_2)-i_2(w_1)} - g(t_1)dt_1 \\
\frac{\partial^2\Lambda((BR(i),i),(BR(i),i))}{\partial i_1(w_2)\partial i_2(w_1)} &= 2\int_0^a \frac{g(t_1)-i_2(w_1)}{i_2(w_2)-i_2(w_1)}(t_1-i_1(w_2))dt_1 \\
\frac{\partial^2\Lambda((BR(i),i),(BR(i),i))}{\partial i_1(w_2)\partial i_2(w_2)} &= 2\int_0^a \frac{i_2(w_2)-g(t_1)}{i_2(w_2)-i_2(w_1)}(t_1-i_1(w_2))dt_1
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2\Lambda((BR(i),i),(BR(i),i))}{\partial i_2(w_1)\partial i_2(w_1)} &= 2\int_0^a \frac{g(t_1)-i_2(w_1)}{i_2(w_2)-i_2(w_1)}i_2(w_1) + \frac{i_2(w_2)-g(t_1)}{i_2(w_2)-i_2(w_1)}g(t_1)dt_1 \\
\frac{\partial^2\Lambda((BR(i),i),(BR(i),i))}{\partial i_2(w_1)\partial i_2(w_2)} &= 2\int_0^a (i_2(w_1)-g(t_1))\frac{i_2(w_2)-g(t_1)}{i_2(w_2)-i_2(w_1)}dt_1
\end{aligned}$$

$$\frac{\partial^2\Lambda((BR(i),i),(BR(i),i))}{\partial i_2(w_2)\partial i_2(w_2)} = 2\int_0^a b - \frac{(i_2(w_2)-g(t_1))^2}{i_2(w_2)-i_2(w_1)} - g(t_1)dt_1$$

Diagonal Language: (only if $b = a$)

$$i_1(w_1) = 2\frac{a}{3},\ i_2(w_1) = \frac{a}{3},\ i_1(w_2) = \frac{a}{3},\ i_2(w_2) = 2\frac{a}{3},\ g(t_1) = t_1$$

$$\nabla^1 = \begin{pmatrix} 2\int_0^a t_1(2\frac{a}{3}-t_1)dt_1 \\ 2\int_0^a t_1(t_1-\frac{a}{3})dt_1 + (2\frac{a}{3}-a)a^2 \\ \int_0^a(2\frac{a}{3}-t_1)t_1dt_1 \\ \int_0^a t_1(t_1-2\frac{2}{3}a)dt_1 + (2\frac{2}{3}a-a)a^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\nabla^2 = \frac{a^2}{3}\begin{pmatrix} 1 & 1 & -1 & 2 \\ 1 & 1 & 2 & -1 \\ -1 & 2 & 1 & 1 \\ 2 & -1 & 1 & 1 \end{pmatrix}$$

The Eigenvalues are $^{-}a^2, \frac{1}{3}a^2\ a^2,\ a^2$ which are not all positive.

Horizontal Language:

$$i_1(w_1) = i_1(w_2) = \frac{a}{2},\ i_2(w_1) = \frac{b}{4},\ i_2(w_2) = 3\frac{b}{4},\ g(t_1) = \frac{b}{2}$$

$$\nabla^1 = \begin{pmatrix} \int_0^a b(\frac{a}{2} - t_1)dt_1 \\ \int_0^a b(t_1 - \frac{a}{2})dt_1 \\ \int_0^a (2\frac{b}{4} - \frac{b}{2})\frac{1}{2}dt_1 \\ \int_0^a \frac{b}{2}(\frac{b}{2} - 2 \cdot 3\frac{b}{4})dt_1 + (2 \cdot 3\frac{b}{4} - b)ab \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

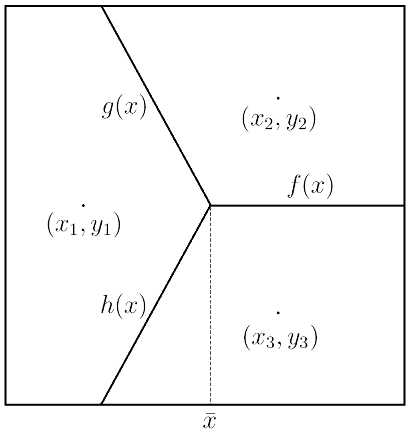
$$\nabla^2 = \begin{pmatrix} \frac{3b^2-a^2}{3b}a & \frac{a^3}{3b} & 0 & 0 \\ \frac{a^3}{3b} & \frac{3b^2-a^2}{3b}a & 0 & 0 \\ 0 & 0 & \frac{3}{4}ab & -\frac{1}{4}ab \\ 0 & 0 & -\frac{1}{4}ab & \frac{3}{4}ab \end{pmatrix}$$

The Eigenvalues are $\frac{ab}{2},\ \frac{3b^2-2a^2}{3b}a,\ ba,\ ba$, which are all positive if $3b^2 > 2a^2$.

A.7 Languages with three words

We derive four equilibria with three words for the unit square type space with uniformly distributed types. The equilibrium conditions are that i) each two interpretations are equidistant from the line that separates their categories and that ii) each interpretation is the gravity point of its category. When solving for an equilibrium, we directly assume the sender strategy $\omega : [0, 1] \to \{\omega_1, \omega_2, \omega_3\}$ to be the best response to the interpretations which implies condition i). Condition ii) is then equivalent to the first derivative of the loss function being equal to zero.

A.7.1 The stable three word language



For $i(w_j) = (x_j, y_j)$ and $w(t)$ the unique best reply to $i(\cdot)$. Then, parametrized to $\{x_j, y_j\}_{j=1,2,3}$, the expected loss can be expressed as

$$\begin{aligned}
& L(x_1, x_2, x_3, y_1, y_2, y_3) \\
= & \int_0^{\bar{x}} \int_{h(x)}^{g(x)} (x_1 - x)^2 + (y_1 - y)^2 dydx \\
+ & \int_0^{\bar{x}} \int_{g(x)}^{1} (x_2 - x)^2 + (y_2 - y)^2 dydx + \int_{\bar{x}}^{1} \int_{f(x)}^{1} (x_2 - x)^2 + (y_2 - y)^2 dydx \\
+ & \int_0^{\bar{x}} \int_{0}^{h(x)} (x_3 - x)^2 + (y_3 + y)^2 dydx + \int_{\bar{x}}^{1} \int_{0}^{f(x)} (x_3 - x)^2 + (y_3 - y)^2 dydx \ ,
\end{aligned}$$

where the separating functions $g(\cdot)$, $f(\cdot)$, and $h(\cdot)$ are defined by

$$\begin{aligned}
f(x) &= \frac{x_3^2 - x_2^2 + y_3^2 - y_2^2}{2(y_3 - y_2)} - \frac{x_3 - x_2}{y_3 - y_2} \cdot x \\
g(x) &= \min\left\{ \frac{x_2^2 - x_1^2 + y_2^1 - y_1^2}{2(y_2 - y_1)} - \frac{x_2 - x_1}{y_2 - y_1} \cdot x, 1 \right\} \\
h(x) &= \max\left\{ \frac{x_1^2 - x_3^2 + y_1^1 - y_3^2}{2(y_1 - y_3)} - \frac{x_1 - x_3}{y_1 - y_3} \cdot x, 0 \right\}
\end{aligned}$$

and $\bar{x}$ is the value of x that solves $g(x) = h(x) = f(x)$:

$$\bar{x} = \frac{1}{2} \frac{(x_2^2 - x_1^2 + y_2^2 - y_1^2)(y_1 - y_3) - (x_1^2 - x_3^2 + y_1^2 - y_3^2)(y_2 - y_1)}{(x_2 - x_1)(y_1 - y_3) - (x_1 - x_3)(y_2 - y_1)}$$

The equilibrium point (solving $\nabla L(x_1^*, x_2^*, x_3^*, y_1^*, y_2^*, y_3^*) = 0$) is

$$\begin{aligned}
(x_1, y_1)^* &= (0.1962024, 0.5)\\
(x_2, y_2)^* &= (0.6827004, 0.7684006)\\
(x_3, y_3)^* &= (0.6827004, 0.2315994)
\end{aligned}$$

This equilibrium point is a local minimum as the Hessian matrix is positive definite:

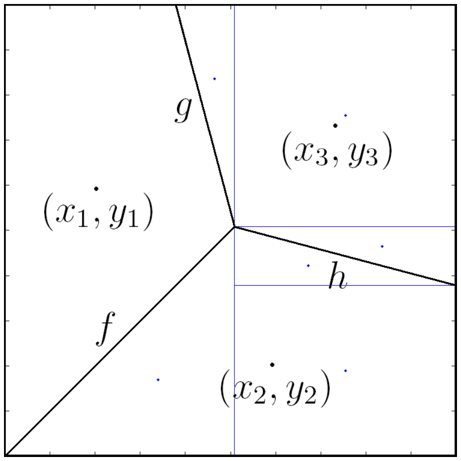
$$H = \begin{pmatrix}
0.5928064 & -0.1002008 & -0.1002008 & 0 & -0.0304094 & 0.0304094\\
-0.1002008 & 0.3718096 & 0.0456908 & -0.1814553 & 0.0480322 & 0.0360234\\
-0.1002008 & 0.0456908 & 0.3718096 & 0.1814553 & -0.0360234 & -0.0480322\\
0 & -0.1814553 & 0.1814553 & 0.4085451 & 0.0333674 & 0.0333674\\
-0.0304094 & 0.0480322 & -0.0360234 & 0.0333674 & 0.4503365 & -0.1305797\\
0.0304094 & 0.0360234 & -0.0480322 & 0.0333674 & -0.1305797 & 0.4503365
\end{pmatrix}$$

with Eigenvalues

$$0.0713266,\ 0.3380698,\ 0.3546698,\ 0.5639263,\ 0.6284244,\ 0.6892269$$

which are all positive.

A.7.2 An unstable three word language



$$\begin{aligned}
f(x) &= \frac{x_1^2 - x_2^2 + y_1^2 - y_2^2}{2(y_1 - y_2)} - \frac{x_1 - x_2}{y_1 - y_2} \cdot x \\
g(x) &= \frac{x_3^2 - x_1^2 + y_3^2 - y_1^2}{2(y_3 - y_1)} - \frac{x_3 - x_1}{y_3 - y_1} \cdot x \\
h(x) &= \frac{x_3^2 - x_2^2 + y_3^2 - y_2^2}{2(y_3 - y_2)} - \frac{x_3 - x_2}{y_3 - y_2} \cdot x
\end{aligned}$$

The loss function is given by

$$\begin{aligned}
& L(i, w_i) \\
= \;& \int_0^{\bar{x}} \int_0^{f(x)} (x - x_2)^2 + (y - y_2)^2 dydx + \int_{\bar{x}}^1 \int_0^{h(x)} (x - x_2)^2 + (y - y_2)^2 dydx \\
+ \;& \int_0^{\hat{x}} \int_{f(x)}^1 (x - x_1)^2 + (y - y_1)^2 dydx + \int_{\hat{x}}^{\bar{x}} \int_{f(x)}^{g(x)} (x - x_1)^2 + (y - y_1)^2 dydx \\
+ \;& \int_{\hat{x}}^{\bar{x}} \int_{g(x)}^1 (x - x_3)^2 + (y - y_3)^2 dydx + \int_{\bar{x}}^1 \int_{h(x)}^1 (x - x_3)^2 + (y - y_3)^2 dydx
\end{aligned}$$

which can be simplified to

$$\begin{aligned}
& L(i, w_i) \\
= \;& (y_2 - y_1) \int_0^{\bar{x}} f^2(x) dx + (y_2 - y_3) \int_{\bar{x}}^1 h^2(x) dx + (y_1 - y_3) \int_{\hat{x}}^{\bar{x}} g^2(x) dx \\
+ \;& \frac{1}{3} + \int_0^{\hat{x}} (x - x_1)^2 - y_1 + y_1^2 dx + \int_{\hat{x}}^1 (x - x_3)^2 - y_3 + y_3^2 dx
\end{aligned}$$

We solve for i^* by using symmetry and the fact that $i^*(w_2)$ is the gravity point of $w^{-1}(i^*(w_2))$ and $i^*(w_3)$ is the gravity point of $w^{-1}(i^*(w_3))$:

(x_2, y_2) being the gravity point of the lower right area implies

$$(x_2, y_2) = \frac{(\frac{2}{3}\bar{x}, \frac{1}{3}\bar{x},)\frac{\bar{x}^2}{2} + (\frac{2}{3}\bar{x} + \frac{1}{3}, \frac{2}{3}h(1) + \frac{1}{3}\bar{x})\frac{(\bar{x}-h(1))(1-\bar{x})}{2} + (\frac{1+\bar{x}}{2}, \frac{h(1)}{2})(1 - \bar{x})h(1)}{\frac{\bar{x}^2}{2} + \frac{(\bar{x}-h(1))(1-\bar{x})}{2} + (1 - \bar{x})h(1)}$$

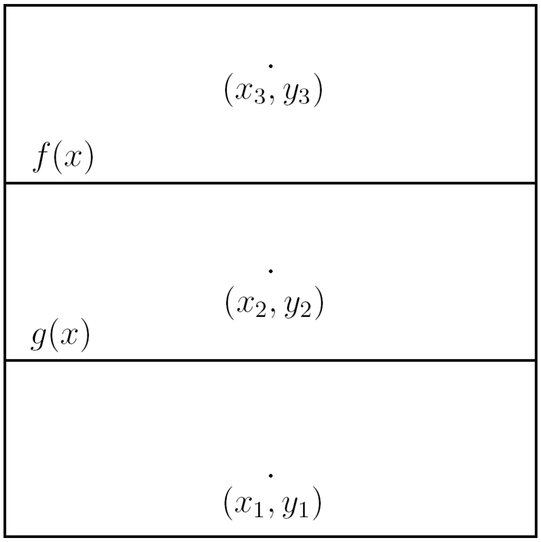
and (x_3, y_3) being the gravity point of the upper right area implies

$$x_3 = \frac{(\frac{1}{3}h(1) + \frac{2}{3}\bar{x})\frac{(\bar{x}-h(1))(1-\bar{x})}{2} + \frac{1+\bar{x}}{2}(1 - \bar{x})^2 + (\frac{2}{3} + \frac{1}{3}\bar{x})\frac{(\bar{x}-h(1))(1-\bar{x})}{2}}{\frac{(\bar{x}-h(1))(1-\bar{x})}{2} + (1 - \bar{x})^2 + \frac{(\bar{x}-h(1))(1-\bar{x})}{2}}$$

The solution to these conditions is given by

$(x_1, y_1) = (0.2033, 0.5921)$ $(x_2, y_2) = (0.5921, 0.2033)$ $(x_3, y_3) = (0.7327, 0.7327)$

A.7.3 The unstable horizontal three word language



$$\begin{aligned}
& L(x_1, x_2, x_3, y_1, y_2, y_3) \\
= \quad & \int_0^1 \int_0^{g(x)} (x - x_1)^2 + (y - y_1)^2 dydx \\
+ \quad & \int_0^1 \int_{g(x)}^{f(x)} (x - x_2)^2 + (y - y_2)^2 dydx \\
+ \quad & \int_0^1 \int_{f(x)}^{1} (x - x_3)^2 + (y - y_3)^2 dydx \\
= \quad & (y_1 - y_2) \int_0^1 g^2(x)dx + (y_2 - y_3) \int_0^1 f^2(x)dx + \frac{1}{3} - x_3 + x_3^2 + \frac{1}{3} - y_3 + y_3^2
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial L(x_1,x_2,x_3,y_1,y_2,y_3)}{\partial x_1} &= 2\int_0^1 (x_1-x)g(x)dx\\
\frac{\partial L(x_1,x_2,x_3,y_1,y_2,y_3)}{\partial x_2} &= 2\int_0^1 (x-x_2)g(x)dx + 2\int_0^1 (x_2-x)f(x)dx\\
\frac{\partial L(x_1,x_2,x_3,y_1,y_2,y_3)}{\partial x_3} &= 2\int_0^1 (x-x_3)f(x)dx - 1 + 2x_3\\
\frac{\partial L(x_1,x_2,x_3,y_1,y_2,y_3)}{\partial y_1} &= \int_0^1 (2y_1-g(x))g(x)dx\\
\frac{\partial L(x_1,x_2,x_3,y_1,y_2,y_3)}{\partial y_2} &= \int_0^1 g(x)(g(x)-2y_2)dx - \int_0^1 f(x)(f(x)-2y_2)dx\\
\frac{\partial L(x_1,x_2,x_3,y_1,y_2,y_3)}{\partial y_3} &= \int_0^1 f(x)(f(x)-2y_3)dx - 1 + 2y_3
\end{aligned}
$$

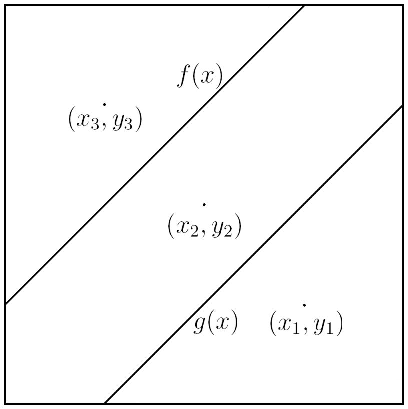
For $(x_1^*, y_1^*), (x_2^*, y_3^*), (x_3^*, y_3^*)$:

$$
H_{(x^*,y^*)} = \frac{1}{6}\begin{pmatrix}
1 & 3 & 0 & 0 & 0 & 0\\
3 & -2 & 3 & 0 & 0 & 0\\
0 & 3 & 1 & 0 & 0 & 0\\
0 & 0 & 0 & 3 & -1 & 0\\
0 & 0 & 0 & -1 & 2 & -1\\
0 & 0 & 0 & 0 & -1 & 3
\end{pmatrix}
$$

With Eigenvalues $-\frac{5}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{2}, \frac{2}{3}, \frac{2}{3}$

As the first Eigenvalue is negative, we conclude that the horizontal language is an unstable saddle point.

A.7.4 The diagonal language



$L = .0964$

$$H = \frac{1}{16}\begin{pmatrix} 3 & 3 & 0 & -3 & 6 & 0 \\ 3 & 2 & 3 & 6 & -6 & 6 \\ 0 & 3 & 3 & 0 & 6 & -3 \\ -3 & 6 & 0 & 3 & 3 & 0 \\ 6 & -6 & 6 & 3 & 2 & 3 \\ 0 & 6 & -3 & 0 & 3 & 3 \end{pmatrix}$$

$$E = -0.9303, 0, 0.1651, \frac{3}{8}, 0.6803, 0.7099$$

Again, one Eigenvalue is negative, therefore the diagonal language is an unstable saddle point.

A.8 Source Code

The source code below runs with Matlab and implements the algorithm described in section 6.2

```
N=100; %number of words
T=100; %total sample size -> approx T/N in each tile
x = rand(1,N); %initial words - x coordinate
y = rand(1,N); %initial words - y coordinate

lambda = 0.95; %inertia (next period = lambda of last period
+ (1-lambda) of current period)
```

```
F=1000; %iterations

min = 2; %initial minimal distance
number = 0; %initial number of samples in each tile
xold = zeros(1,N);
yold = zeros(1,N);

for f = 1:F
   v = rand(T,3); %first column x, second column y, third
column index of closest interpretation.
   v(:,3) = 1;

   for s=1:T
      min = 2;
      for t=1:N;
         if ((v(s,1)-x(t))^2+(v(s,2)-y(t))^2 < min)
            v(s,3) = t;
            min = (v(s,1)-x(t))^2+(v(s,2)-y(t))^2;
         end
      end
   end
   %count vectors close to code
   for n = 1:N
      number = 0;
      for t = 1:T
         if (v(t,3) == n)
            number = number +1;
         end
      end
      %and construct new code
      if number > 0
         xold = x;
         yold = y;
         x(n) = 0;
         y(n) = 0;
         for t = 1:T
            if (v(t,3) == n)
               x(n) = x(n) + v(t,1);
               y(n) = y(n) + v(t,2);
            end
         end
         x(n) = xold(n)*lambda + (1-lambda)*x(n) / number;
         y(n) = yold(n)*lambda + (1-lambda)*y(n) / number;
      end
   end
   %draw figure
   if mod(f,50) == 0
      [f,F]
      if N>2
         voronoi(x,y)
         box on
         axis([0 1 0 1])
      else
         plot(x,y,’.’)
         box on
         axis([0 1 0 1])
      end
      pause(.01);
   end
end
```

References

AGRANOV, M., AND A. SCHOTTER (2008): "Ambiguity and Vagueness in the Announcement (Bernanke) Game: an Experimental Study of Natural Language," .

APALOO, J. (1997): "Revisiting Strategic Models of Evolution: The Concept of Neighborhood Invader Strategies," *Theoretical Population Biology*, 52, 71–77.

AZRIELI, Y. (2009): "Characterization of multidimensional spatial models of elections with a valence dimension," Working Paper.

AZRIELI, Y., AND E. LEHRER (2007): "Categorization generated by extended prototypes - An axiomatic approach," *Journal of Mathematical Psychology*, 51, 14–28.

BHATIA, N. P., AND G. P. SZEGŐ (1970): *Stability Theory of Dynamical Systems*. Springer.

BISHOP, C. M. (2006): *Pattern Recognition and Machine Learning.* Springer.

BLUME, A., Y.-G. KIM, AND J. SOBEL (1993): "Evolutionary Stability in Games of Communication," *Games and Economic Behavior*, 5, 547–575.

BOMZE, I., AND B. PÖTSCHER (1989): *Game Theoretic Foundations of Evolutionary Stability.* Springer Verlag, Berlin.

CRAWFORD, V. P., AND J. SOBEL (1982): "Strategic Information Transmission," *Econometrica*, 50, 1431–1451.

CRESSMAN, R., J. HOFBAUER, AND F. RIEDEL (2006): "Stability of the Replicator Equation for a Single-Species with a Multi-Dimensional Continuous Trait Space," *Journal of Theoretical Biology*, 239, 273–288.

ESHEL, I. (1983): "Evolutionary and Continuous Stability," *Journal of Theoretical Biology*, 103, 99–111.

ESHEL, I., AND E. SANSONE (2003): "Evolutionary and Dynamic Stability in Continuous Population Games," *Journal of Mathematical Biology*, 46, 445–459.

FRYER, R., AND M. O. JACKSON (2008): "A Categorical Model of Cognition and Biased Decision-Making," *The B.E. Press Journal of Theoretical Economics*, 8(1).

GARDENFORS, P. (2000): *Conceptual Spaces: The Geometry of Thought.* MIT Press.

HEIFETZ, A., C. SHANNON, AND Y. SPIEGEL (2007): "What to maximize if you must," *Journal of Economic Theory*, 133(1), 31–57.

HOFBAUER, J., J. OECHSSLER, AND F. RIEDEL (2009): "Brown–von Neumann–Nash Dynamics: The Continuous Strategy Case," *Games and Economic Behavior*, 65, 406–429.

JÄGER, G. (2007): "The evolution of convex categories," *Linguistics and Philosophy*, 30(5), 551–564.

JÄGER, G., AND R. VAN ROOIJ (2007): "Language Stucture: Psychological and Social Constraints," *Synthese*, 159(1), 99–130.

KREPS, D. M., AND R. WILSON (1982): "Sequential Equilibria," *Econometrica*, 50, 863–894.

MATSUI, A. (1991): "Cheap Talk and Cooperation in a Society," *Journal of Economic Theory*, 54, 245–58.

MAYNARD SMITH, J. (1974): "The Theory of Games and the Evolutiona of Animal Conflicts," *Journal of Theoretical Biology*, 47, 209–221.

OECHSSLER, J., AND F. RIEDEL (2001): "Evolutionary Dynamics on Infinite Strategy Spaces," *Economic Theory*, 7, 141–162.

——— (2002): "On the Dynamic Foundation of Evolutionary Stability in Continuous Models," *Journal of Economic Theory*, 107, 223–252.

OKABE, A., B. BOOTS, AND K. SUGIHARA (1992): *Spatial tessellations: concepts and applications of Voronoi diagrams.* Wiley, Chichester.

RITZBERGER, K., AND J. W. WEIBULL (1995): "Evolutionary Selection in Normal-Form Games," *Econometrica*, 63(6), 1371–1399.

Robson, A. (1990): “Efficiency in Evolutionary Games: Darwin, Nash, and the Secret Handshake,” *Journal of Theoretical Biology*, 144, 379–96.

Schlag, K. (1993): “Cheap Talk and Evolutionary Dynamics,” Discussion Paper B-242, Bonn University.

Sobel, J. (1993): “Evolutionary Stability and Efficiency,” *Economics Letters*, 42, 313–319.

Spence, M. (1973): “Job Market Signaling,” *Quarterly Journal of Economics*, 87, 355–374.

Trapa, P., and M. Nowak (2000): “Nash equilibria for an evolutionary language game,” *Journal of Mathematical Biology*, 41, 172–188.

Vickers, G., and C. Cannings (1987): “On the Definition of an Evolutionary Stable Strategy,” *Journal of Theoretical Biology*, 129, 349–353.

Wärneryd, K. (1993): “Cheap talk, coordination and evolutionary stability,” *Games and Economic Behavior*, 5, 532–546.