Linguist and Philos (2007) 30:551–564
DOI 10.1007/s10988-008-9024-3

RESEARCH ARTICLE

The evolution of convex categories

Gerhard Jäger

Published online: 26 March 2008

Abstract Gärdenfors (*Conceptual spaces*, 2000) argues that the semantic domains that natural language deals with have a geometrical structure. He gives evidence that simple natural language adjectives usually denote natural properties, where a natural property is a **convex** region of such a "conceptual space." In this paper I will show that this feature of natural categories need not be stipulated as basic. In fact, it can be shown to be the result of evolutionary dynamics of communicative strategies under very general assumptions.

Keywords Language evolution · Evolutionary game theory · Conceptual spaces · Cognitive semantics

1 Introduction

Model theoretic semantics assumes little or no constraints on the space of possible meanings that go beyond purely formal, set-theoretic properties. For instance, the denotation of an intersective adjective is standardly considered to be any relation between indices and entities, a sentence denotation could be any set of indices, etc. More substantive constraints have been proposed for higher order types like generalized quantifiers (like the requirements of conservativity, etc.), but these constraints can be expressed completely in terms of set theory as well.

Cognitive semanticists, on the other hand, insist that the space of possible meanings is much more structured and constrained. Many authors assume that

G. Jäger (✉)
Faculty of Linguistics and Literature, University of Bielefeld,
PF 10 01 31, 33615 Bielefeld, Germany
e-mail: Gerhard.Jaeger@uni-bielefeld.de

meanings are syntactically structured entities of some kind of formal language. A comparatively lean ontology is proposed by Gärdenfors (2000). He takes it that meanings are arranged in a geometric structure, in so-called *conceptual spaces*. The dimensions of theses spaces are grounded in perception and cognition. The metric over such a conceptual space is related to the cognitive similarity between stimuli.

According to Gärdenfors (2000), a property is a subset of a conceptual space. *Natural properties* are convex regions of such a space. The central thesis of the book is that "most properties expressed by simple words in natural language can be analyzed as natural properties" (pp. 74, 76).

In the remainder of this paper I will argue that languages where meanings are convex regions (of a special kind) are, in a sense to be defined below, optimally adapted to communication. The preference for convex meanings can thus be seen as the result of some process of (cultural) evolution. The basic idea of why evolution leads to convexity has been published before (Jäger and van Rooij 2007). The present paper generalizes this idea to mixed strategies and asymptotically stable sets (rather than just points). To this end, the formal framework of *evolutionary game theory* is employed.

The next two sections present a simple game theoretic model of communication about conceptual spaces. Sections 4 and 5 review basic results on dynamic stability in evolving systems and the significance of game theoretic equilibrium notions for analyzing dynamic stability. Section 6 introduces the central concept of the *Voronoi tesselation* of a space. The main result of the paper is derived in Sect. 7, and Sect. 8 contains concluding remarks.

2 Evolutionary game theory

The concept of evolution originates from biology, but it is applicable in a wide range of domains, including culture and cognition. Quite generally, an evolutionary dynamics ensues in a population of entities (organisms, neuronal patterns, interaction situations etc.) if this population is characterized by the features of *replication, variation*, and *selection*. The nature of the replication mechanism is inessential here; it may be based on DNA copying as in biology, on imitation in the cultural sphere, or on memory as in cognition. It is essential though that certain features are passed on unchanged under replication. As a prerequisite for evolution in the broad sense, the members of the population must vary with respect to such heritable features, and the replicative success of an item must be correlated with these heritable features. If these conditions are met, selection, i.e., the spread of successful variants in the population, will be the consequence.

The expected number of offspring of an individual, given its endowment with heritable characteristics, is called *fitness*. This parameter might depend on the overall composition of the population. To take an example from the cultural sphere, whether a loanword will be accepted in the language community (and thus have a high fitness) does, among other things, depend on whether this word has a synonym in the borrowing language or not. We speak of *frequency dependent selection* if the fitness of a given type of individual depends on the relative frequency distribution of types within the population as a whole.

Evolutionary game theory was developed by theoretical biologists to model this kind of frequency dependent selection in a mathematically precise way (see for instance the standard textbook Hofbauer and Sigmund 1998). Even though originating from biology, this framework is readily applicable to various forms of cultural evolution as well, as witnessed by a wide range of applications in economics, and a growing interest in other behavioral and social sciences like political science or linguistics.

3 Signaling games

Game theory in general antedates evolutionary game theory, and it is generally a framework for the study of strategic interaction of all kinds. In his book *Convention* (Lewis 1969), David Lewis gave a game theoretic model of communication called *signaling game*. The basic idea is rather simple: There are two players, S (the sender) and R (the receiver). At the beginning of the game, a certain meaning m is picked at random and shown to S (but not to R). After that, S can choose a signal f that is transmitted to R. On the basis of his observation of f, R decides for a certain action a. Both S and R have certain preferences over the possible actions of R, and these preferences are conditioned by the meaning m. In a particularly simple version of signaling games, the task of R is simply to guess the meaning m. If his guess is correct, both S and R score one point, otherwise neither of them gets a point. We are thus dealing with a *partnership game*, i.e., a game where the interests of the two players completely coincide. Also, signaling games are *asymmetric games* because the two roles of sender and receiver are not interchangeable.

The *strategies* of both players are functions from observations to actions. So for S, a strategy is a mapping from meanings to signals, and for R the reverse, i.e., a function from signals to meanings.

The sets of possible meanings and possible signals are usually assumed to be both finite. This is unrealistic for communication with natural language as both the set of possible meanings and of possible signals are infinite. I will focus on communication with simple words here though, so the set of signals can in fact be considered to be finite. The set of possible meanings is obviously infinite, however. For instance, each natural number is a possible denotation of some word. Arguably, the set of meanings is actually not just a discrete but a continuous infinity. This is most obvious in the spatial and temporal domain, which are isomorphic to some n-dimensional Euclidean space. Gärdenfors (2000) actually argues that information is cognitively represented in *conceptual spaces*, i.e., geometrically structured sets. Some of these sets are discrete, but some (like spatial and temporal information, but also information about colors or tastes) are continuous. If Gärdenfors' assumptions on the structure of the meaning space are correct, the set of meanings is not just infinite but continuous.

However, the cognitive representation of continuous domains like space or colors has a bounded resolution. Differences can only be perceived and represented as such if the distance between two stimuli exceeds a certain threshold. Likewise, only finite subsets of discretely infinite domains—like the natural numbers—are cognitively represented. This taken into account, the set of meanings is arguably finite after all. However, not every perceivable shade of color or every perceivable stretch of time etc. can be named. It is safe to assume that the cardinality of the set of meanings, while finite, exceeds the size of the set of (morphologically simple) signals by several orders of magnitude.

Under these conditions, communication cannot be guaranteed to be perfect because the number of signals provides an upper bound to the number of meanings that can be communicated in a precise and unambiguous way. So almost all meanings cannot be communicated precisely. Still, there are different degrees of failure in communication. The players should still strive to maximize the *similarity* between the meaning that S wants to express and the interpretation that R assigns to the transmitted signal.

For the remainder of this paper, I assume that the meanings space has an Euclidean structure. The distance between two meanings in this space is inversely related to their similarity. So the preferences of the players are to minimize the distance between intended and guessed interpretation.

In the previous paragraphs I used the language of rationalistic game theory, where conscious players have preferences and pursue their goals. Under the evolutionary interpretation that I assume, these are just metaphors. Strategies are actually behavioral dispositions that can be memorized and imitated. I furthermore assume some kind of positive feedback loop—if a configuration of such dispositions leads to success in communication, their likelihood to be repeated is increased.

To make these assumptions precise, let M, the set of meanings, be a finite subset of some $\mathbb{R}^n$, i.e., an n-dimensional vector space with the usual topology, and F be a finite set of signals. P is a probability distribution over M. Intuitively, $P(m)$ is the probability that the meaning m is to be communicated in a given round of the game. Meanings that are never communicated are irrelevant for the character of the game, so we can assume that $P(m) > 0$ for each m. Finally, let $sim(\cdot,\cdot)$ be a two-place function that measures the similarity between two points in M. It seems reasonable to assume the following properties of this function ($\|x - y\|$ is the Euclidean distance between the points x and y).

$$\begin{gathered}\forall x\colon sim(x,x) = 1\\ \forall x,y\colon sim(x,y) > 0\\ \forall x,y,z\colon \|x-y\| > \|x-z\| \rightarrow sim(x,y) < sim(x,z)\\ \forall x,y,z,w\colon \|x-y\| = \|z-w\| \rightarrow sim(x,y) = sim(z,w)\end{gathered}$$

So we require that the similarity between two points only depends on their distance, it is always a positive number, the maximal similarity is 1, every point is maximally similar to itself but no two different points are maximally similar to themselves, and similarity is inversely related to distance.

A possible strategy S for the sender is a function from meanings to forms, i.e., from M to F. As the domain of such a function is finite, there are finitely many such functions. Likewise, a receiver strategy R is a function from F to M; again there are finitely many such functions.

The utility function of the game is given by

$$u(S,R) = \sum_{m \in M} P(m) sim(m, R(S(m))) \tag{1}$$

Note that we do not distinguish between the utility of the sender and the receiver. Rather, both always obtain the same utility. In other words, the games in question are *partnership games*.

4 The replicator dynamics

Under the evolutionary interpretation of game theory, the utility of a strategy (against a population of strategies) is to be interpreted as the expected number of offspring of an individual playing this strategy. In reality, populations are finite and replication takes place in finite time steps. As the number of offspring of an individual playing a certain strategy is a random variable, the evolution of the strategy profile of a population is necessarily a stochastic process. However, this discrete random process can be approximated by a deterministic continuous time dynamics if populations are sufficiently large, the so-called **replicator** dynamics. The class of games we are dealing with here are *asymmetric games* because there are two distinct roles, sender and receiver, that are not interchangeable. In an evolutionary setting, this amounts to a two-population dynamics, a population of senders and a population of receivers. The replicator dynamics in this case is given by the following system of differential equations:

$$\dot{x}_i = x_i \left(\sum_j y_j \left(u(s_i, r_j) - \sum_k x_k u(s_k, r_j) \right) \right) \tag{2}$$

$$\dot{y}_j = y_j \left(\sum_i x_i \left(u(s_i, r_j) - \sum_k y_k u(s_i, r_k) \right) \right) \tag{3}$$

Here x_i is the relative frequency of the strategy s_i within the population of senders. Likewise, y_j is the relative frequency of the strategy r_j in the receiver population.

In actual communication, each agent can assume either the role of sender or of receiver, depending on context. A perhaps more realistic setting thus assumes just a single population of agents each of which has a disposition both for a certain sender strategy and a receiver strategy. This can be modeled by *symmetrizing* the original asymmetric game. Now a strategy i defines both a sender strategy s_i and a receiver strategy r_i. The utility of two players using the strategy a_i and a_j respectively is the average over the two possible role assignments:

$$u_{sym}(a_i, a_j) = \frac{1}{2}(u(s_i, r_j) + u(s_j, r_i)) \tag{4}$$

The dynamic counterpart of such a symmetric game is a one-population setting which is described by the symmetric replicator dynamics:

$$\dot{z}_i = z_i\left(\sum_j z_j\left(u_{sym}(a_i, a_j) - \sum_k z_k u_{sym}(a_k, a_j)\right)\right) \tag{5}$$

5 Dynamic stability

In most applications of evolutionary game theory, it is impossible or at least impracticable to solve the differential equations of the replicator dynamics analytically. Rather, investigations focus on studying the long term qualitative behavior of an evolutionary system. The notion of *evolutionary stability* is central in this respect. An *evolutionarily stable strategy*[1] (cf. Maynard Smith and Price 1973; Maynard and Smith 1982) is a configuration of a population that (a) is in equilibrium, i.e., the population does not leave that state due to the inherent dynamics, and (b) that is protected against small amounts of mutation. This means that mutations will die out due to the inherent evolutionary dynamics provided the amount of mutants is sufficiently small.

Evolutionarily stable strategies are single strategies. A system which has attained an ESS will remain there, but there is no guarantee that a system will ever attain an ESS. Many games do not even have an ESS. Thomas (1985) generalizes the notion of evolutionary stability to sets of strategies. An *evolutionarily stable set* (ESSet) is a set of strategies that is evolutionarily stable in the sense described above. Each state in such a set is an equilibrium, so a system will remain in this state unless some external perturbation occurs. Furthermore, a sufficiently small amount of mutations will force the system to converge to some equilibrium from the same ESSet (but not necessarily the original equilibrium).

In game theory, an equilibrium state is defined as a *Nash equilibrium*, the single most central concept in modern game theory:

[1] In the sequel, I tacitly assume that strategies can be *mixed*. A mixed strategy is a probability distribution over pure strategies. The utility of a mixed strategy is the expected value of its standard utility.

Definition 1

1. x is a symmetric Nash equilibrium (in a symmetric game) iff

$$\forall y\colon u_{sym}(x,x) \geq u_{sym}(x,y).$$

2. (x,y) is an asymmetric Nash equilibrium (in an asymmetric game) iff

$$\forall z\colon u(x,y) \geq u(z,y)$$

and

$$\forall w\colon u(x,y) \geq u(x,w).$$

Whether or not a set of Nash equilibria is an ESSet depends solely on the utility function of the game. The following definition is adapted from Cressman (2003):

Definition 2 A set E of symmetric Nash equilibria is an *evolutionarily stable set* (ESSet) if, for all $x^* \in E, u(x^*,y) > u(y,y)$ whenever $u(y,x^*) = u(x^*,x^*)$ and $y \notin E$ (Cressman 2003, p. 42).

An ESS is an ESSet which is a singleton.

The notions of ESS and ESSets are static notions that only depend on the utility function and do not make reference to the replicator dynamics (or any other evolutionary dynamics). What we are actually interested in though in this article is the qualitative long-term behavior of populations. However, there is a tight connection between static and dynamic stability concepts for the class of games considered here.

First, the class of signaling games under consideration are *partnership games*. In Akin and Hofbauer (1982) it is shown that for these games, every trajectory converges to some rest point. This means that in partnership games, the population does not enter, e.g., periodic cycles that are known for instance from predator-prey systems. Furthermore, the class of games considered here are symmetrized asymmetric games. Cressman (2003) shows that for this class of games, the ESSets are exactly the *asymptotically stable sets of rest points*. A set is asymptotically stable iff it is stable and attracting. A set is stable iff, loosely speaking, every point within a small topological environment of this set will remain within a small environment. A set is attracting iff every trajectory within some small environment of the set will converge to some point within this set.

Let me briefly explain why this notion is so important. A deterministic dynamics is usually an approximation, while the dynamic behavior of actual observable systems is subject to a certain amount of random noise. If a system is in an asymptotically stable set, it is guaranteed that it will remain within the neighborhood of this set provided this random noise is sufficiently small. On the other hand, the system will eventually leave each set that is not asymptotically stable, either due to the dynamics as such or because of some random perturbation. So in the long run, we can expect that a system will converge to some asymptotically stable set. If a certain feature is shared by all asymptotically stable sets, we can expect this feature to emerge eventually. Under the assumption that some empirically observable system "had enough time" to converge, we can expect to observe that feature. Furthermore, in partnership games the system will always converge to some rest point. We can thus expect that in the long run, the evolutionary dynamics of signaling games will bring the system into some ESSet.

In the remainder of the paper I will show that the fact that the extension of simple expressions is always a convex region of a conceptual space (or rather almost convex, in a sense to be specified below) is a feature that is shared by all elements of some ESSet.

6 Voronoi tesselations and convexity

A set of points is called *convex* if it is closed under the betweenness-relation. As we are dealing with an n-dimensional Euclidean space, a point x is between y and z if x lies on the straight line connecting y and z between y and z. Expressed in terms of analytic geometry, this means that x is a convex combination of y and z, i.e., there is some number α strictly between 0 and 1 such that $x = \alpha y + (1 - \alpha)z$. So we have:

Definition 3 A set E is *convex* iff

$$\forall x,\ y \forall \alpha \in (0, 1) \colon x, y \in E \rightarrow \alpha x + (1 - \alpha)y \in E. \tag{6}$$

As pointed out by Gärdenfors (2000, p. 87), the convexity of concepts can be seen as an epiphenomenon of a more profound generalization. Each concept is associated with a *prototype*, which is a certain point in a conceptual space. (According to prototype theory, this is the point the best exemplifies the concept in question.) If we have n different concepts $c_1, \ldots, c_n$, this means we have n different prototypes $p_1, \ldots, p_n$. We now define the sets E_i (the extensions of the concept c_i) as the set of points that are more similar to p_i than to any other prototype:

$$E_i = \{x | \forall j (\|x - p_i\| < \|x - p_j\|)\} \tag{7}$$

The family of sets $E_1, \ldots, E_n$ is a so-called *Voronoi tessellation* of the space that is generated by the prototypes. (An example of a Voronoi tesselation of a two-dimensional space is given in Fig. 1.)

These sets do not completely exhaust the space because all points which have the same distance to at least two prototypes are excluded. However, *almost all* points of the space belong to some E_i.[2]

It is an elementary fact about Voronoi tessellations (see for instance Aurenhammer 1991) that each cell E_i is a convex set.

[2] The term *almost all* is used in the topological sense here. In this context, almost all points have a property P if and only if the complement set of P has the Lebesgue measure zero.

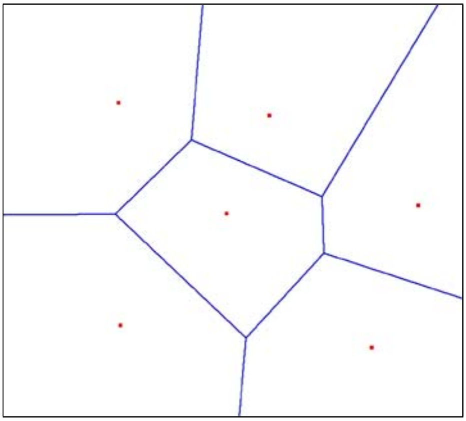

Fig. 1 Example of a Voronoi tesselation

7 Voronoi tessellations and evolutionary stability

Every pure receiver strategy R of the asymmetric game is a mapping from forms to meanings. If there are n forms, the image of R is an m-tuple of meanings, with $m \leq n$. These points in the meanings space can be considered as prototypes $p_1^R, \ldots, p_m^R$. Now let us consider the set *best replies* $BR(R)$ of the sender to R, i.e., the set of strategies S that maximizes $u(S, R)$. By choosing a certain form f to express the meaning m, the sender implicitly chooses a certain prototype $R(f)$ that the receiver will pick out. Since the similarity between the sender's meaning and the receiver's meaning should be maximized, any best reply of the sender to R will express each meaning m by a form f such that $R(f)$ is a prototype that is at least as close to m as any other prototype.

Lemma 1 *For any pure receiver strategy R, it holds that*

$$S \in BR(R) \text{ iff } \forall m\colon S(m) \in R^{-1}(\arg_{p_i} \min \|m - p_i\|). \tag{8}$$

Proof Suppose the right hand side is false. Then there is some m^* such that $S(m^*)$, call it f^*, has the property that $R(f^*) \notin \arg_{p_i} \min \|m^* - p_i\|$. Let S' be like S except that $S'(m^*) = f' \in R^{-1}(\arg_{p_i} \min \|m^* - p_i\|)$. Then $\|m^* - R(S(m^*))\| > \|m^* - R(S'(m^*))\|$, hence

$$sim(m^*, R(S(m^*))) < sim(m^*, R(S'(m^*))),$$

hence $u(S, R) < u(S', R)$, hence $S \notin BR(R)$.

As for the other direction, suppose $S \notin BR(R)$. Then there is some S^* such that $u(S^*, R) > u(S, R)$. This means that for at least one m^*,

$$sim(m^*, R(S^*(m^*))) > sim(m^*, R(S(m^*)))$$

and therefore $\|m^* - R(S^*(m^*))\| < \|m^* - R(S(m^*))\|$. But then the right hand side must be false. □

As an immediate consequence, we get

Corollary 1 *If R is a pure receiver strategy, the inverse image of any $S \in BR(R)$ is consistent with the Voronoi tessellation of the meaning space that is induced by the image of R.*

We say that a partition $F_1, \ldots, F_m$ of a finite set of vectors is consistent with a Voronoi tessellation $E_1, \ldots, E_m$ iff two vectors belong to the same F_i whenever they belong to the same E_j.

Note that the inverse image of any $S \in BR(R)$ possibly *extends* a Voronoi tessellation. This is due to the fact that some meanings may be equidistant to two prototypes, in which case the sender may assign them to either of the two corresponding forms without any impact on the utility. It may thus happen that there are three distinct meanings that are equidistant to the same two prototypes. In this case a best response may assign the first and third meaning to one prototype and the second one to the other. The inverse image of such a sender strategy is not convex then. So we can only say that the partition of the meaning space that is induced by the best reply to some pure receiver strategy is *quasi-convex*, meaning that there is a family of convex sets such that almost all points of the meanings space belong to one of these sets.

How does this relate to evolutionary stability? Let us first consider the dynamic behavior of the asymmetric game. The asymmetric counterpart of the notion of an ESSet is the notion of an *strict equilibrium set* (SESet). The following definition is adapted from Cressman (2003).

Definition 4 *A strict equilibrium set* (SESet) of an asymmetric game is a set F of (possibly mixed) Nash equilibria with the following properties:

- If $(S^*, R^*) \in F$ and $u(S, R^*) = u(S^*, R^*)$, then $(S, R^*) \in F$, and
- if $(S^*, R^*) \in F$ and $u(S^*, R) = u(S^*, R^*)$, then $(S^*, R) \in F$.

(Cressman 2003, p. 70)

So in short, an SESet is a set of Nash equilibria that is closed under mixed-strategy best replies. It is a static characterization of asymptotically stable sets of rest points for the asymmetric replicator dynamics:

Theorem 1 (Cressman) *A set F of rest points of the asymmetric replicator dynamics is asymptotically stable if and only if F is an SESet (Cressman 2003, p. 72).*

An example As pointed out above, the best reply to some *pure* receiver strategy induces an extension of a Voronoi tessellation. This does not necessarily hold for mixed strategies though. To see this, consider the following scenario: The meaning space is the set of multiples of 0.001 within the interval [0, 2]. The similarity function is defined as $\text{sim}(x, y) = \exp(-\|x - y\|^2)$. There are two forms, f_0 and f_1. R is a mixed strategy that maps f_0 to 1, and f_1 with equal probability either to 0 or to 2. The best reply of the sender to R is the unique function that maps all meanings in the interval [0.1712, 1.8287] to f_0, and every

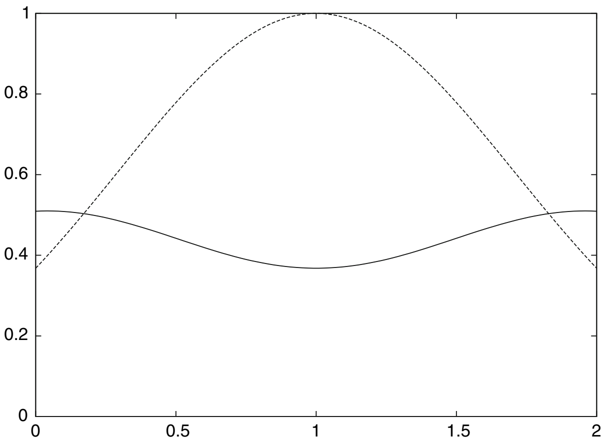

Fig. 2 Utility function for the two sender strategies in the example

other meaning to f_1. Hence the inverse image of S is not consistent with any Voronoi tessellation. Figure 2 contains a plot of the expected utilities in this example. The x-axis displays the meaning space and the y-axis the expected similarity between a sender meaning and a receiver meaning. The dotted line represents the expected similarity if S uses f_0, and the solid line the same for f_1. It is easy to see that at the periphery, f_1 is optimal, while towards the center f_0 is the best option. Obviously, the inverse image of this strategy is not consistent with any Voronoi tessellation.

So the above considerations about best replies always inducing Voronoi tessellations really only apply to best replies to pure receiver strategies. However, it can be shown that in an SESet, every sender strategy is in fact a best reply to some *pure* receiver strategy. This follows from the following theorem:

Theorem 2 (Cressman) *Every SESet is a finite union of Cartesian products of faces of the state spaces (Cressman 2003, p. 71).*

A face of the state space is the set of strategies where at least one pure strategy has a probability 0. So if F is an SESet and $(S, R) \in F$, there are some faces SS and RR with $S \in SS$ and $R \in RR$ such that $SS \times RR \in F$. Each face contains at least one corner, i.e., a pure strategy. Since F is a set of Nash equilibria, S must be a best reply to any pure strategy in RR. Hence every sender strategy that is part on some SESet is a best reply to some pure receiver strategy. Therefore we have

Corollary 2 *If (S, R) is an element of some SESet, the inverse image of S is consistent with some Voronoi tessellation of the meaning space.*

So we know that under the asymmetric replicator dynamics, each state that belongs to some asymptotically stable set of rest points comprises a sender strategy that induces a quasi-convex partition of the state space. How about the symmetrized game? Luckily, there is a tight connection between SESets of some asymmetric game and ESSet of its symmetrization. If strategies of the symmetrized game are simply represented as ordered pairs of a sender strategy and a receiver strategy, this can be expressed as:

Theorem 3 (Cressman) *E is an ESSet of the symmetrization of the asymmetric game G if and only if it is an SESet of G (Cressman 2003, p. 87).*

Hence we obtain, as a direct consequence of the previous results, the following main result of this paper:

Theorem 4 *If a symmetric strategy is an element of some ESSet, the inverse image of its sender strategy is consistent with the Voronoi tessellation that is induced by the image of its receiver strategy.*

To ensure that this is relevant for the dynamics of the system, it remains to be shown that each game in question does have at least one ESSet. It is shown in Hofbauer and Sigmund (1998) that the average fitness of a population is a strict Lyapunov function in partnership games. Since the average fitness is a continuous function of the state of the population and the state space is compact, the fitness has a global maximum. It then follows directly that the set of all states where fitness obtains its global maximum is an asymptotically stable set of rest points, and thus an ESSet.

8 Conclusion

This paper investigated the class of signaling games where the set of meanings is equipped with a Euclidean geometrical structure and where the utility function is not binary—communication is either successful or not—but gradient, being inversely related to the distance between the meaning that the sender tries to communicate and the interpretation that the receiver assigns to the transmitted signal. Such a scenario is particularly interesting in cases where the number of meanings exceeds the number of signals, such that perfect communication is not always possible.

The central question to be addressed was the long term dynamic behavior of a population of agents playing such a game that evolves according to the replicator dynamics. It turned out that in the long run, such a population will evolve towards a state where the sender strategy is consistent with the Voronoi tessellation that is induced by the image of the receiver strategy. An immediate consequence of this is that the sender strategy partitions the meaning space into quasi-convex categories. The only exceptions to the convexity of categories pertain to meanings that have the same distance to two different prototypes.

This result is relevant for cognitive science because the fact that cognitive categories are frequently convex sets in a conceptual space has been noticed before, especially in the work of Peter Gärdenfors.

The main result of this paper can be interpreted in two ways. Under one interpretation, the convexity of cognitive categories is not so much a property of cognition but rather a consequence of a positive feedback loop in communication. To put it succinctly, the explanation is essentially social rather than cognitive.

The notion of a signaling game can also be interpreted in a purely cognitive way though. Under this interpretation, ‘‘meanings’’ in the sense of the model are perceptual stimuli and ‘‘signals’’ are higher-lever representations of categories. A move of a ‘‘sender’’ would be the categorization of some perceptual stimulus, and a move of a receiver the invocation of a prototypical exemplar of a given category. Under this interpretation, the convexity of categories is simply the consequence of a tendency to maximize the similarity of the instances of the same category. Of course these two interpretations, the communicative and the cognitive one, are not mutually exclusive and may both be partially true.

In this paper I assumed that the meaning space has a certain granularity and is therefore finite. This is justifiable because there are limits to the resolution of perceptual stimuli. Also, this decision has a pragmatic aspect because the theorems about the relation between static and dynamic stability concepts used above are only applicable for games with finite strategy spaces.

The dynamic stability of games with infinite strategy spaces is far less understood, and it is currently a topic of active foundational research in economics and theoretical biology (see for instance Metz et al. 1996 or Oechssler and Riedel 2001). It remains to be seen whether modeling the meaning space as a continuous set changes the qualitative predictions of the model. Another important open question is how the probability distribution over meanings affects the structure of the ESSets of the game. The expectation would be that a biased probability distribution will lead to small evolutionarily stable sets where frequent meanings are prototypes. This has in fact been confirmed for simple examples by computer simulations, but so far no general analytical results have been obtained.

References

Akin, E., & Hofbauer, J. (1982). Recurrence of the unfit. *Mathematical Biosciences, 61*, 51–62.

Aurenhammer, F. (1991). Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR), 23*(3), 345–405.

Cressman, R. (2003). *Evolutionary dynamics and extensive form games*. Cambridge (MA): MIT Press.

Gärdenfors, P. (2000). *Conceptual spaces*. Cambridge, MA: The MIT Press.

Hofbauer, J., & Sigmund, K. (1998). *Evolutionary games and population dynamics*. Cambridge: Cambridge University Press.

Jäger, G., & van Rooij, R. (2007). Language structure: Psychological and social constraints. *Synthese, 159*(1), 99–130.

Lewis, D. (1969). *Convention*. Cambridge: Harvard University Press.

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge: Cambridge University Press.

Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature, 246*(5427), 15–18.

Metz, J. A. J., Geritz, S. A. H., Meszéna, G., Jacobs, F. J. A., & van Heerwaarden, J. S. (1996). Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. In

S. J. van Strien & S. M. V. Lunel (Eds.), *Stochastic and spatial structures of dynamical systems* (pp. 183–231). Amsterdam: North Holland.

Oechssler, J., & Riedel, F. (2001). Evolutionary dynamics on infinite strategy spaces. *Economic Theory, 17*(1), 141–162.

Thomas, B. (1985). On evolutionarily stable sets. *Journal of Mathematical Biology, 22*, 105–115.