Cooperation, Conceptual Spaces and the Evolution of Semantics

Peter Gärdenfors[1]and Massimo Warglien[2]

[1] Lund University Cognitive Science, Kungshuset, S-22222 Lund, Sweden
Peter.Gardenfors@lucs.lu.se
[2] Department of Business Economics, Ca' Foscari Università, Venezia, Italy
warglien@unive.it

Abstract. We start by providing an evolutionary scenario for the emergence of semantics. It is argued that the evolution of anticipatory cognition and theory of mind in the hominids opened up for cooperation about future goals. This cooperation requires symbolic communication. The meanings of the symbols are established via a "meeting of minds." The concepts in the minds of communicating individuals are modelled as convex regions in conceptual spaces. We then outline a mathematical framework based on fixpoints in continuous mappings between conceptual spaces that can be used to model such a semantics.

1 Communication and Meaning: An Evolutionary Perspective

When communication first appears, it is the communicative *act* in itself and the context it occurs in that is most important, not the expressive form of the act [1]. As a consequence, the *pragmatic* aspects of language are the most fundamental from an evolutionary point of view. When communicative acts (later speech acts) in due time become more varied and eventually conventionalized and their contents become detached from the immediate context, one can start analyzing the different *meanings* of the acts. Then *semantic* considerations become salient. Finally, when linguistic communication becomes even more conventionalized and combinatorially richer, certain markers, a.k.a. *syntax*, are used to disambiguate the communicative contents when the context is not sufficient to do so. Thus syntax is required only for the subtlest aspects of communication – pragmatic and semantic features are more fundamental.

This view on the evolutionary order of different linguistic functions stands in sharp contrast to mainstream contemporary linguistics. For followers of the Chomskian school, syntax is the primary study object of linguistics; semantic features are added when grammar is not enough; and pragmatics is a wastebasket for what is left over (context, deixis, etc). However, we believe that when the goal is to develop a theory of the evolution of communication, the converse order – pragmatics before semantics before syntax – is more appropriate. In other words, there is much to find out about the evolution of communication, before we can understand the evolution of semantics, let alone syntax.

In support of the position that pragmatics is evolutionarily primary, we want to point out that most human cognitive functions had been chiselled out by evolution

P. Vogt et al. (Eds.): EELC 2006, LNAI 4211, pp. 16 – 30, 2006.

before the advent of language. We submit that language would not be possible without all these cognitive capacities, in particular having a theory of mind and being able to represent future goals (see [2]). This position is not uncontested. Some researchers argue that human thinking cannot exist in its full sense without language (e.g. [3]). According to this view, the emergence of language is a cause of certain forms of thinking, e.g. concept formation.

However, seeing language as a cause of human thinking is like seeing money as a cause of human economics [4, p. 94]. Humans have been trading goods as long as they have existed. But when a monetary system does emerge, it makes economic transactions more efficient and far reaching. The same applies to language: hominids have been communicating long before they had a language, but language makes the exchange of meaning more effective. The analogy carries further. When money is introduced in a society, a relatively stable system of *prices* emerges. Similarly, when linguistic communication develops, individuals will come to share a relatively stable system of *meanings*, i.e. components in their mental spaces, which communicators can exchange between each other. In this way, language fosters a *common structure* of the mental spaces of the individuals in a society.

Within traditional philosophy of language, a semantics is seen as a mapping between a language and the world. From an evolutionary perspective, this view has severe problems. For one thing, it does not involve the users of the language. In particular, it does not tell us anything about how individual user can “grasp” the meanings determined by such a mapping [5]. In this article, we want to propose a radically different view of the evolution of semantics based on a “meeting of minds.” According to this view, the meanings of expressions do not reside in the world or solely in the mental schemes of individual users, but they emerge from the communicative interactions between the language users.

The first part of this paper (sections 2 and 3) presents an evolutionary scenario for the emergence of a “socio-cognitive” semantics. We shall argue that the evolution of anticipatory cognition and theory of mind in the hominids opened up for cooperation about future goals. This cooperation requires symbolic communication. The meanings of the symbols are established via a “meeting of minds.” In the second part of the paper (sections 4-6), we outline a mathematical framework based on fixpoints in continuous mappings between conceptual spaces that can be used to model such a semantics.

This view on how meanings are established gains additional support from a different direction. In a variety of computer simulations and robotic experiments (e.g. [6], [7], [8], [9], [10], [11], [12]), it has been shown that a stable communicative system can emerge as a result of iterated interactions between artificial agents, even though there is nobody who determines any ”rules” for the communication. A general finding of the experiments is that the greater number of ”signallers” and ”recipients” involved in communication about the same outer world, the stronger is the convergence of the reference of the messages that are used and the faster the convergence is attained. Still, different ”dialects” in the simulated community often emerge. However, the “mental spaces” that have been used for robots in these simulations have, in general, been very simplistic and assumed to be identical in structure for all individuals.

2 Cooperation for Future Goals

Language is the solution to certain problems concerning communication. But animals communicate without language. So what are the communicative reasons for developing a more complicated system like human symbolic language? Our answer is that humans have a capacity to communicate about their future goals.

To elaborate this position, we must analyze some of the cognitive prerequisites for symbolic language. Bischof [13] and Bischof-Köhler [14] argue that animals other than humans cannot anticipate future needs or drive states. Their cognition is therefore bound to their present motivational state (see also [15]). This hypothesis, which is called the Bischof-Köhler hypothesis [16], is supported by the current evidence concerning planning in non-human animals.

Gulz [15] calls planning for present needs *immediate planning* while planning for the future is called *anticipatory planning*. Humans can predict that they will be hungry tomorrow and save some food, and we can imagine that the winter will be cold, so we start building a shelter already in the summer. There is nothing in the available evidence concerning animal planning, notwithstanding all its methodological problems, which suggests that any other genus than *Homo* can represent their *future* desires (the recent results by Mulcahy and Call [17] are not really counterevidence to the thesis). The cognition of other animals concerns here and now, while humans are mentally both here and in the future.

Anticipatory planning is a component in a more general anticipatory cognition that is a hallmark of *Homo sapiens* [18]. It also includes episodic memory [19] and other aspects of "mental time travel" [16], [20]. A central question is what factors along the hominid line have created selective evolutionary forces that have resulted in anticipatory cognition in general (including episodic memory) and anticipatory planning in particular (also cf. [21]).

One answer is provided by Osvath and Gärdenfors [18], who argue that the Oldowan culture led to the co-evolution of transport and anticipatory planning. The hominid life on the savannah during the Oldowan era opened up for many new forms of cooperation for future goals. For example, Plummer [22, p. 139] writes: ”Given that body size often predicts rank in the carnivore guild, an individual *Homo habilis* would likely not have fared well in a contest with many of its contemporary carnivores. Competition with large carnivores may have favoured cohesive groups and coordinated group movements in *Homo habilis*, cooperative behaviour including group defence, diurnal foraging (as many large predators preferentially hunt at night) with both hunting and scavenging being practiced as the opportunities arose, and the ability (using stone tools) to rapidly dismember large carcasses so as to minimize time spent at death sites."

For most forms of cooperation among animals, it seems that mental representations are not needed. If the common goal is *present* in the actual environment, for example food to be eaten or an antagonist to be fought, the collaborators need not focus on a joint representation of it before acting. If, on the other hand, the goal is distant in time or space, then a *mutual* representation of it must be produced before cooperative action can be taken. For example, building a common dwelling requires coordinated planning of how to obtain the building material and advanced collaboration in the construction. In general terms, cooperation about future goals requires that *the mental spaces of the individuals be coordinated* (or, in some cases, negotiated).

3 The Need for Symbols in Communication About Future Goals

Symbolic language is the primary tool by which humans can make their inner worlds known to each other. In previous work [2], [18], [23], [24], it has been proposed that there is a strong connection between the evolution of anticipatory cognition and the evolution of symbolic communication. In brief, the argument is that symbolic language makes it possible to *cooperate about future goals* in an effective way.

Language is based on the use of representations as stand-ins for entities, actual or just imagined. Use of such representations replaces the use of environmental cues in communication. If somebody has an idea about a goal she wishes to reach, she can use language to communicate her thoughts. In this way, language makes it possible for us to coordinate our visions about the future – our minds can meet. The question that has to be answered is why symbolic communication is necessary for this kind of communication.

Tomasello [25, p. 95] defines symbolic communication as the process by which "one individual attempts to manipulate the attention of, or to share attention with, another individual. In specifically linguistic communication [...] this attempt quite often involves both (a) reference, or inviting the other to share attention to some outside entity (broadly construed), and (b) predication, or directing the other's attention to some currently *unshared* features or aspects of that entity [...]." We cannot fully accept this definition. One aspect that is missing in his characterization is that depending on the character of the "outside entity," different cognitive demands on the individual whose attention is manipulated will be relevant. To understand the differences, one must distinguish between (1) entities that are present in the shared environment, (2) entities that are not present in time or space but about which there is some common knowledge, and (3) entities that are unknown to the other individual. Communication about future goals often involves entities of the third kind.

Depending on which type of entity is communicated about, different minimal forms of communication are required. It becomes very natural to map the three kinds of entities to be communicated about to Peirce's [26] triad of index, icon and symbol:

(1) If the entity is present, then *indexical* communication, for example pointing, is sufficient. Animal communication consists almost exclusively of signals, referring to what is present at the moment in the environment, be it food, danger or a mate. This form of communication does not presume that the signaller ascribes any mental representation of the communicated object in the mind of the receiver. It is important to note that this kind of communication does not require any form of symbolic communication. (This is another reason we do not fully accept Tomasello's definition presented above.) Consequently, as long as all communication concerns present entities, there will be no evolutionary pressures for the use of symbols.

(2) If the communicated entity is not present, direct signalling will not work. If I want to refer to a deer that I saw down by the riverside yesterday, merely pointing will not help, nor will a call signal. This form of communication clearly requires representations that are detached from the present [2]. *Iconic miming* may establish the reference, but only if the signaller and receiver have sufficient *common knowledge* about the indicated entity and there are sufficient cues from previous communication or the environment to make it possible for the receiver to identify the object. (This would be a case of what is called triadic miming in [27]. When the relevant entity is

an action, this form of communication works particularly well. By using icons, one agent can show another how to act in order for the two of them to reach a common goal. Icons can work as imperatives, urging the agents to "Do like this!" [23].

(3) The most difficult type of communication concerns *novel* entities that do not yet exist. Collaboration about future, non-existent goals falls within this category. Here the signaller can neither rely on common knowledge about the entity, nor on cues from the environment. Iconic communication might work in exceptional cases, but we submit that it is for this kind of communication that *symbols* prove their mettle. For example, if I have come up with an idea about how to build a new kind of defence wall around our camp, it is very difficult to see how this can be communicated by miming alone. In particular, if the communication involves the *predication* of Tomasello's definition above, that is, directing the other's attention to some currently unshared features or aspects of that entity, symbols seem to be crucial. Such a predication process will also require the productivity and compositionality of a symbolic system.

In this characterization we use "symbolic communication" in a basically Peircian way, meaning that the act is conventional and breaks up compositionally into meaningful sub-acts that relate systematically to each other and to other similar acts [27], [28]. This form of communication is, as far as we know, uniquely human. In this context it should be noted that Tomasello's [25, p. 95] definition of symbolic communication that was presented above also covers what we call indexical and iconic cases.

An important feature of the use of symbols in cooperation is that they can set the cooperators free from the goals that are available in the present environment. The future goals and the means to reach them are picked out and *shared* through the symbolic communication. This kind of sharing gives humans an enormous advantage concerning cooperation in comparison to other species. We view this advantage as a strong evolutionary force behind the emergence of symbols. More precisely, we submit that there has been a co-evolution of cooperation about future goals and symbolic communication (cf. the "ratchet effect" discussed in [4], pp. 37-40 and [18]. However, without the presence of anticipatory cognition, the selective pressures that resulted in symbolic communication would not have emerged. However, once symbolic communication about future goals has been established, it can be used for other purposes, for example, sharing myths and rituals.

We want to show that this kind of sharing mental representations leads to the emergence of a semantics, that is, a set of shared meanings. In our opinion, semantics can be seen as conventionalized pragmatics [29]. One important question then concerns how the cognitive structure of the semantic conventions looks like. Here, so called cognitive semantics offers a cue to one part of the answer (e.g. [29], [30], [31]). According to cognitive semantics, the meanings of words can be represented as "image schemas" in the heads of the communicators. These schemas are abstract mental pictures having an inherent spatial structure constructed from elementary topological and geometrical structures like "container," "link" and "source-path-goal." A common assumption is that such schemas constitute the representational form that is common to perception, memory, and semantic meaning.

However, a general problem for such a semantic theory is: if everybody has their own mental space, how can we then talk about a representation being *the* meaning of an expression? In other words, how can individual mental representations become conventions? Therefore, the question in focus will be: how can language help us *share* our mental spaces?

In the computer simulations and robotic experiments performed by Steels and others, the typical communicative situation is a ”guessing game” [8] where the signaller, by uttering an expression, tries to make the recipient identify a particular object in the environment. It should be noted that in such guessing games (as in Wittgenstein's language games), the participants are only concerned with finding the appropriate referent among those that are present on the scene. In contrast, communication about *non-present* referents, which are in focus here, demands that the communicators have more advanced representational capacities.

4 Semantics as a Meeting of Minds

Our view on the evolution of symbolic communication puts novel, non-present and even fictitious referents in focus. Therefore, a semantic theory that starts from reference to the world seems unnatural from our perspective. Our task is to develop a semantic theory that fits with the evolutionary account presented above. In our view, the semantics does not consist of a mapping from linguistic expressions to an external world, but is rather constituted of the individuals’ mental spaces and mappings between them. In brief, we see semantics as a meeting of the minds and hence we advocate a form of socio-cognitive semantics.

As a comparison, consider the “cognitive semantics” where image schemas have been core carriers of meaning. An image schema is a conceptual structure that belongs to a particular individual. However, when the authors within cognitive linguistics write about image schemas, they are often presented as structures that are *common* to all speakers of a language. However, in the socio-cognitive type of semantics we model in the next section, we do not assume that everybody has the same meaning space, but only that there exist well-behaved mappings between the meaning spaces of different individual – “well-behaved” in the sense that the mappings have certain mathematical properties (to be specified in the following section). As we shall argue, semantic equilibria can exist without assuming shared spaces. The semantics will be represented by a *fixpoint* in the mapping between individual mental spaces.

The fundamental role of communication is to affect the states of mind of others. A meeting of the minds means that the representations in the minds of the communicators will become sufficiently compatible so that successful joint action can arise. Thus we conceive of semantics as a *product of communication* – meanings arise as a result of communicative interactions. The mental space that generates the meanings for a particular individual is partly determined from the individual’s interaction with the world, partly from her interaction with others and partly from her interaction with herself (e.g. in the form of self-reflection). This view does not entail that different individuals mean the same thing by using an expression, only that their communication is sufficiently successful.

As a preparation for our analysis of communication about novel and non-present objects as a basis for semantics, let us consider a theoretical scenario proposed by Freyd [32]. The main theme of her paper is that knowledge, by the fact that it is *shared* in a language community, imposes *constraints* on individual cognitive representations. She argues that the structural properties of individuals' mental spaces have evolved because "they provide for the most efficient sharing of concepts," and proposes that a dimensional structure with a small number of values on each dimension will be especially "shareable." This process of creating shared meanings is continually ongoing: the interplay between individual and social structures is in eternal co-evolution. The effects are magnified when communication takes place between many individuals (cf. the simulations by Steels and others).

The constraints of sharing concepts can be discussed in relation to the image schemas of cognitive semantics. Even if different individuals do not have identical schemas, there are good reasons to assume that they have developed a high degree of similarity. One is that since basic image schemas are supposed to represent perceptual and other bodily experiences, the very fact that humans have similar constitutions makes it likely that our representations are very similar. Another reason is that if the image schema corresponding to a particular expression is markedly different for two individuals, it is likely that this will lead to problems of communication. A desire for successful communication will therefore lead to a gradual alignment among the members of a linguistic community of the image schemas.

Image schemas provide a bridge between a focus on shared meanings and a focus on the common shape of underlying conceptual structures that facilitate mutual understanding and the successful interaction between possibly different but similarly structured mental spaces. After all, we can communicate effectively even if we do not share the same mental representation. For example, in communication between children and adults, children often represent their concepts using fewer dimensions, and dimensions that are different from those of the adults.

Our aim is to model how a common structure in individual mental spaces will ensure the existence of a "meeting of minds," and how semantics may be grounded in the formal properties of such interaction.

5 Meeting of Minds as Fixpoints in Communication Games

In this section we outline, in rather broad terms, a mathematical framework for semantics as "meeting of the minds".

As long as communication is conceived as a process through which the mental state of an individual affect the mental state of another one, a "meeting of the minds" is a condition in which both individuals find themselves in compatible states of mind that do not require further processing. Just like covenants shake hands after reaching an agreement on the terms of a contract, speakers may reach a point in which both believe they have understood what they are talking about. Of course, they may actually mean different things, just like the terms of a contract might prove to be interpreted differently by the covenants. But it is enough that, in a given moment and a given context, speakers may reach a point in which they feel there is a mutual understanding – no matter whether mutual agreement implies or not that they mean the same thing.

A very common mathematical way to define such kind of state would be to identify it as a fixpoint. A fixpoint x* of a function f(x) is a point in which the function maps x* on itself (f(x*) = x*). But what kind of object is a function that reaches a fixpoint when minds agree? The most natural candidate for such a semantics is a function that maps language expressions on mental states, and vice-versa – a kind of interpretation function and its inverse. So, in our framework minds meet when the interpretation function mapping states of mind on states of mind via language finds a resting point – a fixpoint.

Using fixpoints is, of course, not new to semantics. The semantics of programming languages often resort to fixpoints to define the "meaning" of a program: its meaning is where the program will stop (for a remarkable review, see [33]). In a different vein, Kripke's [34] theory of truth is grounded on the notion of a fixpoint – in his case the fixpoints of a semantic evaluation function are at the focus of his interest. Fixpoints are also crucial in other fields, such as the study of semantic memory: content-addressable memories usually store information as a fixpoint of a memory update process (the canonical example being [35]).

However, here we make a fairly different use of the fixpoint notion to define our "meeting of minds" semantics, since we consider the fixpoints of an interactive, social process of meaning construction and evaluation. From this point of view, our use of fixpoints resembles more the one made by game theorists to define states of mutual compatibility of individual strategies. To some extent, we are following the tradition of communication games ([36], [37], etc), but to this tradition we are adding some assumptions about the *topological* and *geometric* structure of the individual mental spaces that will allow us to specify more substantially how the semantic emerges and what properties it has.

Our argument is that some types of topological and geometric properties of mental representations are more likely to engender meetings of minds, because they lend more naturally fixpoints to communication activities. Thus, we shift from the conventional emphasis on the way we share the same concepts to an emphasis on the way the "shape" of our conceptual structures makes it possible for us to find a point of convergence. A parallel with the pragmatics of conversation in the Gricean tradition comes to the mind: just like maxims of conversation ensure that talk exchanges find a mutually accepted direction, we explore the complementary notion that the way we shape our concepts deeply affects the effectiveness of communication.

On this ground, we make an implicit selection argument: just like wheels are round because they make transportation efficient, we expect to identify the shapes of concepts that are selected to make communication smooth.

It turns out that structural properties of conceptual representations that grant the existence of meetings of minds are to a large extent already familiar to cognitive semantics and in particular to the theory of conceptual spaces. These basic properties are the metric structure induced by similarity, the closed/bounded nature of concepts, convexity of conceptual representation, and the assumption that natural language, with all its resources, can "translate" (spatial) mental representations with reasonable approximation. In what follows, we will make more precise these notions and the role they play in a "meeting of minds" semantics theory.

Our first step is to assume, following [38], that conceptual spaces are made out of primitive *quality dimensions* (often rooted in sensorial experience) and that similarity provides the basic metric structure to such spaces. The dimensions represent various "qualities" (colour, shape, weight, size, position …) of objects in different domains.

While the nature of psychologically sound similarity measures is still highly controversial (and presumably differs between domains), numerous studies suggest that it is a continuous function of Euclidean distance in the conceptual spaces. Consequently, we will assume, as a first approximation, that conceptual spaces can be modelled as *Euclidean spaces*.

It is not only in humans that one finds these kinds of representations. For example, Gallistel [39] devotes an entire chapter to "Vector spaces in the nervous system." in his book on learning mechanisms in biological systems. He writes [39, p. 477]: "The purpose of this chapter is to review neurophysiological data supporting the hypothesis that the nervous system does in fact quite generally employ vectors to represent properties of both proximal and distal stimuli. The values of these representational vectors are physically expressed by the locations of neural activity in anatomical spaces of whose dimensions correspond to descriptive dimensions of the stimulus." Furthermore, it is well known that even fairly simple neural processing mechanisms can approximate arbitrary continuous functions [40].

In [38], it is proposed that *concepts* can be modelled as *convex regions* of a conceptual space. While convexity may seem a strong assumption, it is a remarkably regular property of many conceptual representations grounded in perception (e.g., colour, taste, pitch). Furthermore, we will soon argue that convexity is crucial for assuring the effectiveness of communication.

There are interesting connections between analyzing concepts as convex regions and the *prototype theory* developed by Rosch and her collaborators (see, for example, [30], [41], [42], [43]). When concepts are defined as convex regions of a conceptual space, prototype effects are indeed to be expected. In a convex region one can describe positions as being more or less central. In particular, in a Euclidean space one can calculate the centre of gravity of a region.

It is possible to argue in the converse direction too and show that if prototype theory is adopted, then the representation of concepts as convex regions is to be expected. Assume that some quality dimensions of a conceptual space are given, for example the dimensions of colour space, and that we want to decompose it into a number of categories, for example colour concepts. If we start from a set of prototypes p_1, ..., p_n of the concepts, for example the focal colours, then these should be the central points in the concepts they represent. The information about prototypes can be used to generate concepts by stipulating that p belongs to the same concept as the *closest* prototype p_i. It can be shown that this rule will generate a decomposition of the space – the so-called *Voronoi tessellation*. An illustration of the Voronoi tessellation is given in figure 1.

A crucial property of the Voronoi tessellation of a conceptual space is that it always results in a decomposition of the space into *convex* regions (see [44]). In this way, the Voronoi tessellation provides a constructive geometric answer to how a similarity measure together with a set of prototypes determine a set of categories.

As long as concepts are closed and bounded regions of conceptual spaces, they acquire one more crucial topological property: *compactness*. Euclidean metrics, compactness and convexity set the stage for our fixpoint argument. But before getting there, a last point must be made briefly. A basic tenet of cognitive semantics is that language can preserve the spatial structure of concepts. One way to say it is that

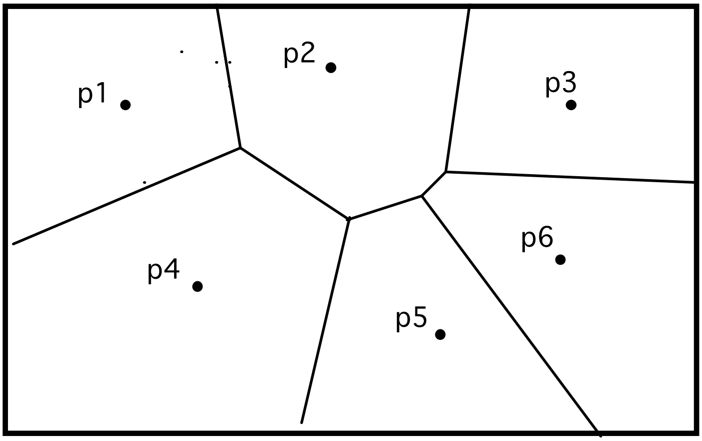

Fig. 1. Voronoi tessellation of the plane into convex sets

language can preserve the neighbourhood relations among points of conceptual spaces. In topology, a neighbourhood preserving function is nothing but a continuous function. In other words, assuming that language can preserve neighbourhood relations of conceptual spaces implies assuming that language can establish a continuous mapping between mental spaces of different individuals – and a continuous mapping of the product space of individual mental spaces on itself. While this continuity assumption may seem extreme, it basically says that natural language must have enough plasticity to map neighbourhoods of points in a conceptual space on neighbourhoods of points in another conceptual space – or in the space itself. Furthermore, as we shall see, this assumption can be relaxed to assume that such continuous mappings can be suitably approximated.

Now all ingredients are there, and we can simply remind you of one of the most fundamental results of analysis, Brouwer's [45] theorem: each continuous map of a convex, compact set on itself has at least one fixpoint. In the present context, this result basically tells us that, no matter what is the content of individual mental representations, provided that such representations are "well shaped" and that language is plastic enough to preserve the spatial structure of concepts, there will always be at least one point representing a "meeting of minds." Furthermore, given a continuous function and convex compact spaces, whenever such spaces can be decomposed in smaller convex closed subsets (they can be "triangulated"), there will always be a function mapping such decomposition on itself (called a "simplicial approximation") that will approximate the continuous function, and preserve its fixpoint property. In other words, such fixpoints may be still approximated by a coarser mapping.

After this short and very informal mathematical detour, our central claim should become apparent: whenever the facility to reach a meeting of minds matters, convex mental representations provide a background over which language can deploy most of its power. We see this as an indirect explanation of why concepts are in general convex. Please note that we are not claiming that convex representations are "faithful" representations of the world – we just claim that since they are effective, one should find them quite widespread. In fact, our claim implies that one should expect to find convex representations even in cases in which they are biased representations of the world: seeing a non convex world with convex spectacles might be a peculiar bias arising from selective pressures towards effective communication.

Brouwer's theorem provides us with an existence result that guarantees that an appropriate meeting of minds can be found among a set of communicators that have convex and compact mental representations of meaning. However, the result does not in itself say very much about the contents of the fixpoint or how it can be reached.

6 Two Examples

We now proceed to briefly present two examples of how the meeting of minds can emerge in communicative systems. There are many ways in which such an interactive semantics can be established. In some cases it can be a proper game of communication, where the meeting of minds can be interpreted as an equilibrium of the game. In other cases, it can result from simpler adaptive processes that do not require strategic reasoning.

Jäger and van Rooij [46] provide an example of the first kind. Their domain is the colour space and the problem they approach is how a common meaning for colour terms can develop in a communication game. In their example, there are only two players: s (signaller) and r (receiver). It is assumed that the two players have a common conceptual space C for colour. Jäger and van Rooij define the space as a "continuous space" but from their following claims, it clearly must be a compact and convex space, such as a colour circle or a colour spindle. There is also a fixed and finite set M of n messages that the signaller can convey to the receiver. The colour space C can also be interpreted as a state space from which Nature draws points according to some continuous distribution p. The signaller can choose a *decomposition* S of the space C in n subsets assigning to each colour a unique message. The receiver can choose where to locate n points, corresponding to the meaning assigned to each of the n messages by the signaller.

The goal of the communication game is to maximize the average similarity between the intention of the signaller and the interpretation of the receiver. The communication game unfolds as follows. Nature chooses some point in the colour space, according to some fixed probability distribution. The signaller s knows the choice of nature, but the receiver r does not. Then s is allowed to send one of the messages to r. The receiver r in turn picks a point in the colour space. In the game, s and r maximize utility if they maximize the similarity between nature's choice and r's "interpretation". Here is it only assumed that the similarity is a monotonically decreasing function of the Euclidean distance in the colour space between nature's choice and r's choice.

A Nash equilibrium of the game is a pair (R*, S*), where R* is an n-tuple of points of C and S* is a decomposition (in n subsets) of C, such that both are a best response to each other. Jäger and van Rooij [46] show how to compute the best response functions for each player. The central result of their paper can be restated by saying that if the colour space is convex and compact and the probability and similarity functions are continuous, then there exists a Nash equilibrium, and it corresponds to a Voronoi tessellation of the colour space (which results in convex subsets).

They also show how, in a simplified evolutionary version of the game, convex colour regions can emerge as the evolutionary stable solutions of the game. Jäger and van Rooij's model is also interesting because it provides an illustration of how a discrete system of signs (there are only n signs in their communication game) can give rise to continuous functions mapping agents' mental representations on themselves. In their example, signs define an array of locations in the colour space, and the "best response function" of s and r continuously maps configurations of such array of points as responses to decompositions of C, and vice versa. In this language game, "language" has to be plastic enough to grant the continuity of the "best response" function, and the meaning space C must have enough topological structure to afford the existence of fixpoints. Language plasticity is given here by the possibility to continuously deform the decomposition S and the location of the points of R.

As a second example, Hutchins [47] provides a case that is more loosely related to a game structure, but where fixpoints with a semantic valence result form simple adaptive dynamics shaped by communication. He models how individuals may reach an agreement over an interpretation of potentially incomplete and noisy signals from the environment. Each individual is represented as a constraint satisfaction network, in which nodes represent features (corresponding to quality dimensions in a conceptual space) of the world and connections between nodes impose some coherence between configurations of features.

Communication between agents is modelled through connections between nodes of different agents. Through such connections the state of mind of an agent affects the states of mind of the other agents by transmitting the activation values of its nodes. In other words, communication continuously maps the state of minds of each agent on the states of mind of other agents in a feature space: Imagine a "feature-based" language through which agents can express their beliefs about the state of the world.

Hutchins shows by simulations how agents starting form different beliefs can converge towards fixpoints that represent consensual interpretations of the state of the world. Consensus needs not to correspond to "reality": In many cases it is a form of groupthink, convergence to beliefs dominated more by peer pressure than truth. Revisiting more formally Hutchins' model, Marchiori and Warglien [48] prove that communication can give rise to new fixpoints that were not contained in individual initial memories – i.e. there may be genuine new meanings arising as meetings of minds among communicating agents.

7 Conclusion

In this article, we have first told a story about the evolution of communication based on the unique human capacity for planning for future goals. A consequence of our story is that in order for communication about non-present objects to succeed, the minds of the interlocutors must meet. In the second part of the paper, we have then presented a framework for how this process can be modelled as a fixpoint semantics. To some extent, we have followed the tradition of communication games, but the most innovative part of our model is the assumptions about the *topological* and *geometric* structure of the mental spaces of the communicators. We have focused on the compactness and convexity of these spaces and, following Gärdenfors' [38] work on conceptual spaces, argued that these assumptions are very natural. These assumptions make it possible for us to apply Brouwer's fixpoint theorem, which in this context, is interpreted as saying that for communicators with "well-behaved" mental spaces, there will always exist a meeting of their minds that represents the meaning of the expressions they use. We have also outlined two examples of how such a meeting can be achieved.

The fixpoint semantics that we have presented provide us with rather new perspectives on the functioning of semantics for natural languages. We hope to develop the model to show that this perspective is fruitful and that it can solve many of the problems for classical forms of semantics, for example problems concerning the reference of expressions for non-existing objects and that it can shed new light on the meaning of metaphors.

References

1. Winter, S.: Expectations and Linguistic Meaning. Lund University Cognitive Studies 71, Lund (1998)
2. Gärdenfors, P.: How Homo Became Sapiens: On the Evolution of Thinking. Oxford University Press, Oxford (2003)
3. Dennett, D.: Consciousness Explained. Little, Brown and Company, Boston (1991)
4. Tomasello, M.: The Cultural Origins of Human Cognition. Harvard Unversity Press, Cambridge, MA (1999)
5. Harnad, S.: The Symbol Grounding Problem. Physica D. 42 (1990) 335-46
6. Hurford, J.: The Evolution of Language and Languages. In: Dunbar, R., Knight, C., Power, C. (eds): The Evolution of Culture. Edinburgh University Press, Edinburgh (1999) 173-193
7. Kirby, S.: Function, Selection and Innateness: The Emergence of Language Univerals. Oxford University Press, Oxford (1999)
8. Steels, L.: The Talking Heads Experiment. Laboratorium, Antwerp (1999)
9. Steels, L.: Social and Cultural Learning in the Evolution of Human Communication. In: Oller, K., Griebel, U. (eds.): The Evolution of Communication Systems. MIT Press, Cambridge, MA (2004) 69-90
10. Kaplan, F.: L'émergence d'un lexique dans une population d'agents autonomes. Ph. D. Thesis. Laboratoire d'Informatique de Paris 6, Paris (2000)
11. Vogt, P.: Bootstrapping Grounded Symbols by Minimal Autonomous Robots. Evol. of Comm. 4 (2000) 89-118
12. Vogt, P.: The Emergence of Compositional Structures in Perceptually Grounded Language Games. Artif. Intell. 167 (2005) 206-242
13. Bischof, N.: On the Phylogeny of Human Morality. In: Stent, G. (ed.): Morality as a Biological Phenomenon. Abako, Berlin (1978) 53-74
14. Bischof-Köhler, D.: Zur Phylogenese menschlicher Motivation. In: Eckensberger, L.H., Lantermann, E.D. (eds.): Emotion und Reflexivität. Urban & Schwarzenberg, Vienna (1985) 3-47
15. Gulz, A.: The Planning of Action as a Cognitive and Biological Phenomenon. Lund University Cognitive Studies 2, Lund (1991)
16. Suddendorf, T., Corballis M.C.: Mental Time Travel and the Evolution of Human Mind. Genetic, Social and General Psychology Monographs 123 (1997) 133-167
17. Mulcahy, N.J., Call, J.: Apes Save Tools for Future Use. Science 312 (2006) 1038-1040.

18. Osvath, M., Gärdenfors, P.: Oldowan Culture and the Evolution of Anticipatory Cognition, Lund University Cognitive Studies 121, Lund (2005)
19. Tulving, E.: How Many Memory Systems are There? Am. Psychologist 40 (1985) 385-398
20. Suddendorf, T., Busby J.: Mental Time Travel in Animals? Trends in Cog. Sci. 7 (2003) 391-396
21. Savage-Rumbaugh, E.S.: Hominin Evolution: Looking to Modern Apes for Clues. In: Quiatt, D., Itani, J. (eds.): Hominin Culture in Primate Perspective. University Press of Colorado, Niwot (1994) 7-49
22. Plummer T.: Flaked Stones and Old Bones: Biological and Cultural Evolution at the Dawn of Technology. Yearbook of Phys. Anthrop. 47 (2004) 118-164
23. Brinck, I., Gärdenfors, P.: Co-operation and Communication in Apes and Humans. Mind and Lang. 18 (2003) 484-501
24. Gärdenfors, P.: Cooperation and the Evolution of Symbolic Communication. In Oller, K., Griebel, U. (eds.): The Evolution of Communication Systems. MIT Press, Cambridge, MA (2004) 237-256
25. Tomasello, M.: On the Different Origins of Symbols and Grammar. In: Christansen, M.H., Kirby, S. (eds.): Language Evolution. Oxford University Press, Oxford (2003) 94-110
26. Peirce, C.S.: The Collected Papers of Charles Saunders Peirce. Vols. 1-4. Harvard University Press, Cambridge, MA (1931-35)
27. Zlatev, J., Persson, T., Gärdenfors, P.: Bodily Mimesis as the "Missing Link" in Human Cognitive Evolution. Lund University Cognitive Studies 121, Lund (2005)
28. Deacon, T.W.: The Symbolic Species. Penguin Books, London (1997)
29. Langacker, R.W.: Foundations of Cognitive Grammar, Vol. 1. Stanford University Press, Stanford, CA (1987)
30. Lakoff, G .: Women, Fire, and Dangerous Things. The University of Chicago Press, Chicago, IL (1987)
31. Talmy, L.: Force Dynamics in Language and Cognition. Cognitive Science 12 (1988) 49-100
32. Freyd, J.: Shareability: The Social Psychology of Epistemology. Cognitive Science 7 (1983) 191-210
33. Fitting, M.: Fixpoint Semantics for Logic Programming: A Survey. Theor. Comput. Sci. 278 (2002), 25-51
34. Kripke, S.: Outline of a Theory of Truth. J. of Phil. 72 (1975) 690-716
35. Hopfield, J.J.: Neural Networks and Physical Systems with Emergent Collective Computational Abilities. Proc. Nat. Acad. of Sci. 79 (1982), 2554–2558
36. Lewis, D.: Convention. Harvard University Press, Cambridge, MA (1969)
37. Stalnaker, R.: Assertion. Syntax and Semantics 9 (1979) 315-332
38. Gärdenfors, P.: Conceptual Spaces: The Geometry of Thought. MIT Press, Cambridge, MA (2000)
39. Gallistel, C. R.: The Organization of Learning. MIT Press, Cambridge, MA (1990)
40. Hornik, K., Stinchombe, H., White, H.: Multilayer Feedforward Networks are Universal Approximators. Neural Networks 2 (1989) 359-366
41. Rosch, E.: Cognitive Representations of Semantic Categories. J. of Exp. Psych.: General 104 (1975) 192–233
42. Rosch, E.: Prototype Classification and Logical Classification: The Two Systems. In: Scholnik, E. (ed.): New Trends in Cognitive Representation: Challenges to Piaget's Theory. Lawrence Erlbaum Associates, Hillsdale, NJ (1978) 73–86
43. Mervis, C., Rosch, E.: Categorization of Natural Objects. Ann. Rev. of Psychol. 32 (1981) 89–115

44. Okabe, A., Boots, B., Sugihara, K.: Spatial Tessellations: Concepts and Applications of Voronoi Diagrams. John Wiley & Sons, New York (1992)
45. Brouwer, L.E.J.: Über ein eindeutige, stetige Transformation von Flächen in sich. Mathematische Annalen 69, blz. (1910) 176-180
46. Jäger, G., van Rooij, R.: Language Structure: Psychological and Social constraints, Synthese (to appear)
47. Hutchins, E.: Cognition in the Wild. MIT Press, Cambridge, MA (1995)
48. Marchiori, D., Warglien, M.: Constructing Shared Interpretations in a Team of Intelligent Agents: The Effects of Communication Intensity and Structure. In: Terano, T., Kita, H., Kaneda, T., Arai, K., Deguchi, H. (eds.): Agent-Based Simulation: From Modeling Methodologies to Real-World Applications. Springer Verlag, Berlin (2005)