Semantics, conceptual spaces, and the meeting of minds — Source link

Massimo Warglien, Peter Gärdenfors

Institutions: Ca' Foscari University of Venice

Related papers:

- Conceptual Spaces: The Geometry of Thought

- The Geometry of Meaning: Semantics Based on Conceptual Spaces

- Language structure: psychological and social constraints

- Cognitive Representations of Semantic Categories.

- Event structure, conceptual spaces and the semantics of verbs

Semantics, conceptual spaces, and the meeting of minds

Massimo Warglien · Peter Gärdenfors

Abstract We present an account of semantics that is not construed as a mapping of language to the world but rather as a mapping between individual meaning spaces. The meanings of linguistic entities are established via a "meeting of minds." The concepts in the minds of communicating individuals are modeled as convex regions in conceptual spaces. We outline a mathematical framework, based on fixpoints in continuous mappings between conceptual spaces, that can be used to model such a semantics. If concepts are convex, it will in general be possible for interactors to agree on joint meaning even if they start out from different representational spaces. Language is discrete, while mental representations tend to be continuous—posing a seeming paradox. We show that the convexity assumption allows us to address this problem. Using examples, we further show that our approach helps explain the semantic processes involved in the composition of expressions.

M. Warglien
Department of Management, Advanced School of Economics, Università Ca' Foscari Venezia, Venezia, Italy
e-mail: warglien@unive.it

M. Warglien
Laboratory for Experimental Economics / Fondazione Ca' Foscari, Venezia, Italy

P. Gärdenfors (✉)
Department of Philosophy, Box 188, 22100 Lund, Sweden
e-mail: Peter.Gardenfors@lucs.lu.se

🖄 Springer

1 Introduction

Within traditional philosophy of language, semantics is seen as mapping between a language and the world (or several "possible worlds"). 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 users can grasp the meanings determined by such a mapping (Harnad 1990; Gärdenfors 1997). Another tradition, cognitive semantics, brings in the language user by focusing on the relations between linguistic expressions and the user's mental representations of the expressions' meanings, typically in the form of so-called *image schemas*; but cognitive semantics has problems explaining the social nature of semantics.

In this article, we propose a radically different view 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 rather emerge from the communicative interaction of language users. The fundamental role of human communication is, indeed, to affect the state of mind of others, "bringing about cognitive changes" (van Benthem 2008). A meeting of minds means that the representations in the minds of the communicators become sufficiently compatible to satisfy the goals that prompted the communication.

As an example of how such a meeting of minds can be achieved by communication but without the aid of language, consider declarative pointing (Bates 1976; Brinck 2004, Gärdenfors and Warglien, to appear). This consists of one individual pointing to an object or location and, at the same time, checking that the other individual (the "recipient") focuses her attention on the same object or location. The recipient, in turn, must check that the "sender" notices the recipient attending to the right entity. This "attending to each others' attention" is known as *joint attention* (Tomasello 1999) and is a good, though fallible, mechanism for checking that the minds of the interactors meet, by focusing on the same entity.[1]

Achieving joint attention can be seen as reaching a fixpoint in communication. When my picture of what I point out to you agrees with my understanding of what you are attending to, my communicative intent is in equilibrium. Conversely, when what you attend to agrees with your understanding of what I want to point out to you, your understanding is in equilibrium (Gärdenfors and Warglien, to appear).[2] At the other end of the communicative spectrum, a purely symbolic kind of meeting of minds is the agreement of a contract. We will return to such agreements in Sect. 3.

When the interactors are communicating about the external world, pointing is sufficient to make minds meet and agree on a referent. When they need to share referents in their inner mental spaces, they require a more advanced tool. This is where language proves its mettle (Brinck and Gärdenfors 2003; Gärdenfors 2003; Gärdenfors

---

[1] Note that pointing, unlike language, is a continuous way of referring to the outer world: the direction of the finger is continuously variable.

[2] Two features of this process should be noted. First, joint attention requires a so-called *theory of mind* that includes second-order attention, in that both communicative partners can attend to the attention of the other. Second, the two partners need not have the same image of each other's inner state: it is perfectly possible for you and I to reach joint attention without my picture of your attention being aligned with your attention. In other words, joint attention never requires transcending one's own subjective realm.

and Osvath 2010). In a way, language is a tool for reaching joint attention in our inner worlds. Goldin-Meadow (2007, p. 741) goes beyond our metaphorical assertion to write that, in children, "pointing gestures form the platform on which linguistic communication rests and thus lay the groundwork for later language learning."

We will assume that our inner worlds can be modelled as spaces with topological and geometric structure, using conceptual spaces (Gärdenfors 2000) as the main modeling tool. The conceptual spaces that carry the meanings for a particular individual are determined partly 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 approach does not entail that different individuals mean the same thing by the same expression, only that their communication is sufficiently effective.

As a comparison, consider the models of cognitive semantics (see e.g. Lakoff 1987; Langacker 1986, 1987; Croft and Cruse 2004; Evans 2006) where image schemas have traditionally been the core carriers of meaning. An image schema is a conceptual structure belonging to a particular individual. The mathematical structures of image schemas are seldom spelled out.[3] A natural way to do this is with topological and geometric notions.

Image schemas are, in general, presented as structures common to all speakers of a language. Given the socio-cognitive type of semantics we model in this paper, we do not assume that everybody has the same meaning space, only that there exist well-behaved mappings between the meaning spaces of different individuals—well-behaved in the sense that the mappings have certain mathematical properties as specified below.

Our approach takes significant inspiration from the communication games studied by Lewis (1969, 1979), Stalnaker (1979) and others (e.g. Schelling 1960; Clark 1992; Skyrms 1998; Parikh 2000, 2010; Jäger and van Rooij 2007). To this foundation, we add assumptions about the topological and geometric structure of the various individuals' conceptual spaces that allow us to specify more substantially how the semantics emerges and what properties it has. Linguistic acts can best be understood as moves in such games. Not only may the players in a communication game have different payoff functions; they may also have different meaning spaces. As suggested earlier, we will show that semantic equilibria can exist without needing to assume that the communicating individuals possess the same mental spaces.

So long as communication is conceived as a process through which the mental state of one individual affects the mental state of another, then a "meeting of the minds" will be that condition in which both individuals find themselves in compatible states of mind, such that no further processing is required. Just as bargainers shake hands after reaching agreement on the terms of a contract, so speakers reach a point at which both believe they have understood what they are talking about. Of course, they may actually mean different things, just as the bargainers might interpret the terms of the contract differently. It is enough that, in a given moment and context, speakers reach a point at which they *believe* there is mutual understanding. The ubiquity of "assent" signals in conversations (Clark 1992) nicely demonstrates the importance of the mutual awareness of such meetings of mind in everyday communication.

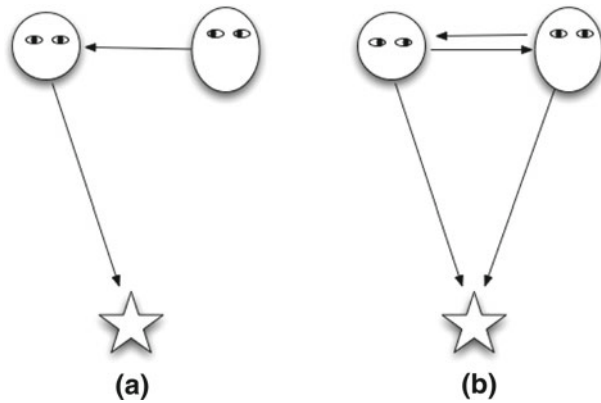[3] For an early computational model, see Holmqvist (1993).

Fig. 1 Joint attention as arriving at a fixpoint

A common way to define such a state mathematically is to identify it as a *fixpoint*. A fixpoint x* of a function $f(x)$ is that point at which the function maps x* onto itself ($f(x^*) = x^*$). What kind of thing is a function that reaches a fixpoint where minds agree? In linguistic communication, the most natural candidate is a function that maps language expressions onto mental states and vice-versa: a kind of interpretation function and its inverse expression function. In our framework, minds meet when the function mapping states of mind onto states of mind, via language, finds a fixpoint.[4]

For a simple example of convergence to a fixpoint as a meeting of minds, consider again the example of joint attention achieved via pointing: e.g., by a child pointing out something to an adult. The relevant mental spaces are, in this case, taken to be each person's visual fields, which may only partially overlap. The goal of the child's pointing is to make the adult react by looking at the desired point in her visual field. The fixpoint is reached when the child sees that the adult's attention is directed at the correct location, and the adult believes her attention is directed to what is being pointed at (see Fig. 1). That fixpoint is characterized by four properties: (1) the attended object is in my line of gaze, (2) the attended object is in your line of gaze, (3) I see that your line of gaze has the right orientation, and (4) you see that my line of gaze has the right orientation. Our lines of gaze intersect at the object, and our representations of each other's gaze are consistent.

Now consider an example involving explicit linguistic interaction. Definite reference, even when verbally performed, generally reflects similar properties. Clark (1992) suggests a collaborative process in which speaker and addressee "work together in the making of a definite reference" (Clark 1992, p. 107). Creating such a reference is a coordination problem that rarely reduces to uttering the right word at the right time. What is required instead is a process of mutual adjustment between speaker and addressee converging on a mutual acceptance that the addressee has understood the

---

[4] Communication is like a dog on a leash pulling its master towards something the dog has in mind. The dog will pull and its master will follow until an equilibrium is reached. The place where they stop is, literally, a fixpoint.

speaker's utterance. The process is highly iterative, involving a series of reciprocal reactions and conversational moves usually concluded by assent signals. Think of the complex series of further requests and information extensions, as well as corrections, nods, and interjections, employed in a simple communicative act such as explaining to a tourist where to find a restaurant she is looking for. Such a process has the clear nature of arriving at a fixpoint. Note that conversational adjustments towards mutual agreement typically resort to both the discrete resources of spoken language and the continuous resources of gesture, intonation, and other bodily signals.

Using fixpoints is, of course, nothing new to semantics. Programming language semantics often resort to fixpoints to define the meaning of a particular program; its meaning is where the program will stop. (For a remarkable review, see Fitting 2002.) In a different vein, Kripke's (1975) theory of truth is grounded on the notion of a fixpoint, but his focus of interest is the fixpoints of a semantic evaluation function. Fixpoints are crucial in other fields, such as the study of semantic memory; content-addressable memories usually store information as the fixpoint of a memory update process. (The canonical example is Hopfield neural nets: see Hopfield 1982.)

We will make a different use of fixpoints to define our "meeting of minds" semantics: namely, as the result of an interactive, social process of meaning construction and evaluation. Our use of fixpoints resembles the attempt made by game theorists to define equilibrium as a state of mutual compatibility among individual strategies'.[5] (See Parikh's (2000) equilibrium semantics for a related approach.) Some topological and geometric properties of mental representations afford meetings of minds because they more naturally generate communicative activity fixpoints. Following Gärdenfors (2000), we will represent concepts as convex regions of mental spaces (see next section).[6] In this way, we shift from a conventional emphasis on the way we share (the same) concepts to an emphasis on the way the shapes of our conceptual structures make it possible to find a point of convergence. A parallel to the Gricean tradition of conversation pragmatics comes to mind. Just as conversational maxims ensure that dialogues follow a mutually acceptable direction, so the way we shape our concepts deeply affects communication's effectiveness.

From this foundation, we make a conjecture about implicit selection: just as wheels are round because they make transportation efficient, conceptual shapes should be selected because they make communication smooth and memorization efficient. Both the convexity and the compactness of concepts play central roles. Constraints on conceptual structure facilitate the creation of coordinated meanings. Communication works so long as it preserves the structure of concepts. This will naturally lead to a consideration of the role of continuous mappings in conveying meaning similarity. In Sect. 3.2 we will elaborate on the important point that similarity preservation can be performed by a discrete system: to wit, the expressions of language. Here we will

---

[5] Note that our fixpoints are of a topological nature, while those used most often in computation and logic exploit properties of monotone functions on (partially) ordered structures.

[6] A region is convex whenever any point lying between two points in the region will itself belong to that region.

focus on noun phrases and indexical expressions; but our approach can be extended to other linguistic categories, particularly verb phrases.[7]

A few points of clarification will be helpful before proceeding to a more formal argument.

2 The topology of conceptual spaces

The structural properties of conceptual representations that make possible the meeting of minds are, to a large extent, already to be found in cognitive semantics and, in particular, the theory of conceptual spaces. These include the metric structure imposed by similarity; the closed (or "bounded") nature of concepts; the convexity of concepts; and the assumption that natural language, with all its resources, can effectively translate mental representations with reasonable approximation. In what follows, we will make these notions, along with the role they play in a "meeting of minds" semantic framework, more precise.

The first step is to assume, following Gärdenfors (2000), that conceptual spaces are constructed out of primitive quality dimensions (often grounded in sensory experience) and that similarity provides the basic metric structure to such spaces. The dimensions represent various qualities of objects (e.g., color, shape, weight, size, force, position) in different domains.[8]

Recall that a metric space is defined by a set of points with a measure of the distance between the points. A *metrizable* topological space is any space whose topological structure can be induced by some suitable metric. Our fundamental assumptions are that conceptual spaces are metrizable and that their metric structure is imposed by a similarity relation. This leaves open the possibility of many different metric structures. While the precise nature of psychologically sound similarity metrics remains highly controversial (and presumably differs between domains), numerous studies (e.g., Shepard 1987; Nosofsky 1988a) suggest it to be a continuous function of Euclidean distance within conceptual spaces. We will assume, as a first approximation, that conceptual spaces can be modelled as Euclidean spaces. However, the general ideas may be applied to other metric structures (see e.g. Johannesson 2002).

Following Gärdenfors (2000), we will define concepts as regions in a conceptual space. Two properties of such regions are worth mentioning. First, so long as concepts are closed and bounded regions in Euclidean conceptual spaces, they acquire a further, critical topological property: compactness.[9] One important intuition regarding compactness is that it provides "enough" points that are near to a set—this proves critical when fixpoints come to be defined. Compactness makes it possible to approximate the entire space through a finite number of points, which will also turn out to be critical to what follows.

---

[7] For some suggestions on how verbs may be analysed, see Gärdenfors (2007) and Gärdenfors and Warglien (to appear b). Different kinds of pointing are analysed in Gärdenfors and Warglien (to appear a).

[8] Note that conceptual spaces are used to represent relations between concepts, a more general usage than the representing of single image schemas in cognitive semantics.

[9] This follows as a corollary of the classic Bolzano-Weierstrass theorem.

Second, as proposed in Gärdenfors (2000), concepts can be modeled as convex regions of a conceptual space. Convexity may seem a strong assumption, but it is a remarkably regular property of many perceptually grounded concepts: e.g., color (Jäger 2007, 2010), taste vowels (2007, 2010). Although the main argument for convexity in Gärdenfors (2000) is that it facilitates the learnability of concepts, it is also crucial for assuring the effectiveness of communication. Although we will in this paper take the concepts of individuals as given, learnability and effectiveness of communication clearly do interact in complementary ways in the acquiring of concepts.

There are interesting comparisons to make between analyzing concepts as convex regions and the prototype theory developed by Rosch and her collaborators. (See for example Rosch 1975, 1978; Mervis and Rosch 1981; Lakoff 1987.) When concepts are defined as convex regions in a conceptual space, prototype effects are to be expected. Given a convex region, one can describe positions in that region as being more or less central. In a Euclidean space, one can also calculate the center of gravity.

Conversely, 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 color space; and that the intention is to decompose it into a number of categories: in this case, color concepts.[10] If one starts from a set of concept prototypes $p_1, \ldots, p_n$—say, the focal colors—then these prototypes should be the central points in the concepts they represent. The information about prototypes can then be used to generate convex regions by stipulating that any point p within the space belongs to the same concept as the closest prototype p_i. This rule will generate a certain decomposition of the space: a so-called Voronoi tessellation. An illustration of a Voronoi tessellation is given in Fig. 2. The particular tessellation is two-dimensional, but Voronoi tessellations can be extended to any arbitrary number of dimensions.

The basic idea is that the most typical meaning of a word or linguistic expression is the prototype at the centre of the convex region assigned to the word.[11] This proves to be a central principle in connecting the continuity of mental spaces with the discreteness of language.

A Voronoi tessellation always results in a decomposition of the space into convex regions. (See Okabe et al. 1992.) The Voronoi tessellation provides an intuitive geometric answer for how a similarity measure together with a set of prototypes can determine a set of categories.

The Voronoi tessellation has a dual graph, the so-called Delaunay triangulation, which will turn out to be useful later on (Okabe et al. 1992). A Delaunay triangulation is obtained by drawing a line segment between the centers (prototypes) of Voronoi cells that share a side or vertex. Barring special cases, this rule will result in a triangulation of the space; even in the special cases, the result can easily be extended to a

---

[10] Since the borders of neighbouring regions simultaneously belong to each region, we call the result a decomposition and not a partitioning. The compactness of the regions is preserved.

[11] In Gärdenfors (2000), it is argued that the region assigned to a linguistic expression may not be constant, but is in general dependent on the context of the use of the expression.
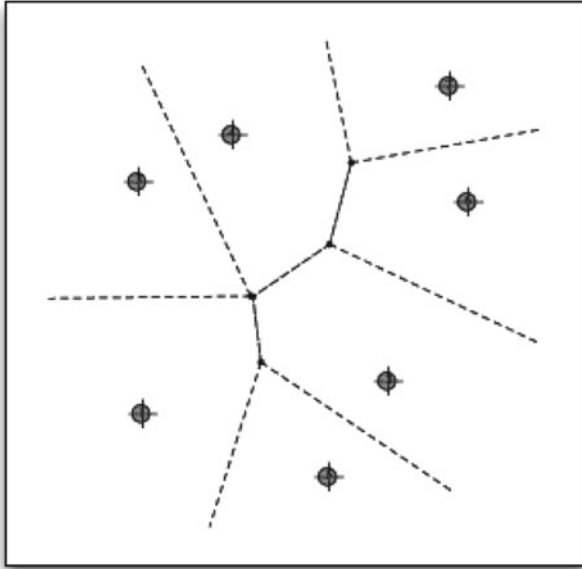
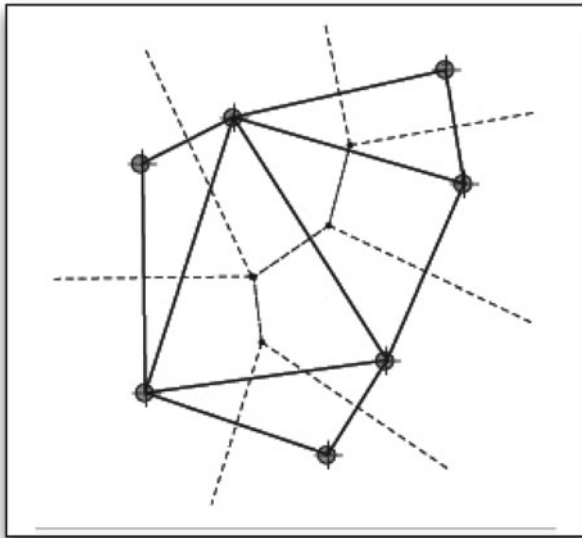**Fig. 2** A Voronoi tessellation of the plane into convex sets

Fig. 3 A Delaunay triangulation (*solid lines*) superimposed on a Voronoi diagram (*dashed lines*)

triangulation. An important property of any Delaunay triangulation is that contiguous prototypes are connected. Such triangulations play a special role in the approximation of continuous functions, as we will show (Fig. 3).

Euclidean metrics, compactness, and convexity together set the stage for our fixpoint argument. Before getting there, one last point must be made, briefly. A basic tenet of cognitive semantics is that language is able to preserve the spatial structure of concepts. One way to express this is that language preserves the nearness relations among points in conceptual spaces. In topology, a nearness-preserving function is nothing other than a *continuous* function. In other words, assuming that language is able to preserve nearness relations in conceptual spaces implies assuming that language is able to establish a continuous mapping between the mental spaces of different individuals—and, as we shall show, a continuous mapping of the product space of an individual's mental space onto itself. While this assumption may seem overly strong, it basically says that natural language must have enough plasticity to map near points in one conceptual space onto near points in another conceptual space. We will show below that this assumption can be relaxed somewhat, to allow that such continuous mappings can be suitably approximated.

3 Fixpoints

3.1 Examples

We have already suggested that meetings of minds can be modelled as fixpoints in some kind of meaning space. We have offered referring by pointing as a prototypical example of finding a communication fixpoint. We next present a few additional examples to illustrate the idea, before providing a more formal treatment of our claim.

Garrod and Anderson (1987) defined an experimental conversational setup in which two agents have to solve the problem of locating each other while moving around in a computer-game labyrinth. They are able to talk with each other. The task amounts to generating descriptions of the maze that are compatible enough to generate appropriate, coordinated action. Dialogue analysis revealed that the agents developed different representations of the task, but that within such diversity, they tended to narrow in on compatible representations: not necessarily identical, but preserving structural relationships such as nearness, direction (right or left), and relative position (below or above). In a similar way, different subway maps can preserve the essential spatial relationships between the stations in a network. The choice of description language seemed to be dependent upon the agent's mental model, with the mapping from representations to language tending to preserve the relationships from the model. It is clear that the choice of terms must preserve the way the elements of the representations are profiled (e.g., rows and columns rather than paths), along with their associated spatial properties: "speakers are adopting particular 'languages' of description in relation to the basic …model, and it is possible to give a reasonable account of these languages if we make a few straightforward assumptions about the form of the underlying model" (Garrod and Anderson 1987, p. 202).

The structure-preserving mappings must work symmetrically in production and interpretation: convergence is facilitated by the interlocutors symmetrically playing the role of hearer and speaker, while keeping the same representations and mappings

in both roles. Convergence results from local adjustments, based on mutual consistency checking of consecutive utterances. Consistency checks can be based on simple heuristics (of which completing the other's sentence may be considered the limit case) that need not imply any overall consideration of common ground or the intentions of the other, only assessing whether one can interpret an utterance in relation to one's own representation of the current circumstances: if "two steps along" makes no sense given one's current position, there is a misalignment. Where consistency checks fail, repair is often attempted by reformulating the utterance in terms that may lead—often do lead, after only a few rounds—to mutual agreement.

Experiments on the emergence of human communication systems often reveal a similar process, whereby mutual understanding emerges as a fixpoint from a joint coordination effort. Some of these experiments use (graphical) sign systems, with the advantage that such systems preserve the continuity of conceptual spaces while verbal systems discretize it—a problem we will analyze later on. For example, Galantucci (2005) created a "virtual environment" (reminiscent of Garrod and Anderson's 1987 setup) consisting of rooms in which players were located without being able to see each other. They had to meet in the same room using a limited number of moves, communicating only through a system of signs on an electronic pad. Sending a sign corresponded to attempting to push the mind of the receiver to locate the right region of the virtual space, while understanding the meaning of the sign received corresponded to mentally locating the correct region. So long as participants were not correctly understanding the signals, the communication system remained out of equilibrium, participants adjusting their reaction to the signs received (and adjusting the signs they sent) until they found a response in which both correctly located the meaning of the sign itself—in this case their reaction to the intended region is the region itself, which is a clear fixpoint.

Other experiments reveal an analogous structure. In another laboratory-based experiment, Selten and Warglien (2007) studied the emergence of a communication system through a referential coordination game, in which players had to establish a coordinated mapping of simple geometric objects and arbitrary sequences of elementary signs. In such a game, mutual understanding corresponds to achieving an equilibrium: in a state of mutual understanding, players' interpretations of the signs support each other, establishing a self-confirming joint mapping of signs and referred-to objects. (Selten and Warglien (2007) go on to explore how such joint mappings can exhibit compositionality, an issue we will consider in Sect. 4.)

In addition to these experimental studies, Jäger and van Rooij (2007) provide a theoretical modeling example of how fixpoints (or, in their case, Nash equilibria) can be used to model a meeting of minds. They further show how the emergence of meaning as the fixpoint of a continuous function can result in the discretization of conceptual spaces. Their choice of domain is the color space; the problem they examine is how a common meaning for color terms can develop through a communication game. In their example, there are only two players: s (the signaler) and r (the receiver). Jäger and van Rooij assume that the two players have a common conceptual color space C. S has a fixed and finite set of n messages that she can convey to the receiver. The color space C can be interpreted as a state space from which nature selects points according to some continuous distribution p. S can choose a decompo-

sition of C into n subsets, assigning each color a unique message. R can choose where to locate n points, corresponding to the meaning assigned to each of the n messages by s.

The goal of the communication game is to maximize the average similarity between the intention of s and the interpretation of r.[12] A Nash equilibrium in the game is a pair (S^*, R^*), where S^* is s's decomposition (into n subsets) of C and R^* is r's n-tuple of points in C, such that both S^* and R^* are a best response to each other. The central conclusion of their paper can be restated by saying that if the color space is convex and compact and if the probability and similarity functions are continuous, then a Nash equilibrium exists, and it corresponds to a Voronoi tessellation of the color space (which results, of course, in convex subsets).

Jäger and van Rooij (2007) model illustrates how a discrete system of signs (there are only n signs in their game) is compatible with a continuous functions, mapping agents' mental representations onto themselves. The signs define an array of locations in the color space; the best-response function of s and r continuously maps configurations of that array as responses to decompositions of C, as well as vice versa. In this game, "language" must be plastic enough to permit the continuity of the best response function, and C must have enough topological structure to afford the existence of fixpoints. Language plasticity is provided by the possibility to continuously deform the decomposition S^* and the location of the points R^*. Notice that adding new signs would involve only local changes in the Voronoi tessellation. In other words, one need not revise all linguistic meanings each time one learns a new word.

3.2 Existence of fixpoints

The examples we have cited suggest a few important issues relating to the semantic use of fixpoints. One, of course, is simply the existence of such fixpoints, and what features of agents' conceptual spaces may favor them. Another concerns how to reconcile the continuous nature of mental spaces with the inherent discreteness of verbal language. A third is how motivation, and the stakes of communication, shape semantic interaction, as well as how fixpoints can be achieved. We will deal in turn with each one of these.

In our pointing example as well as in Jäger and van Rooij's (2007) model, it was assumed that the communicators share more or less the same mental space: in our example, the visual field; in their model, the color space. However, in general, different individuals will have different mental spaces. For simplicity, assume that there are only two individuals A and B with mental spaces C_1 and C_2, both of which are convex and compact. If A communicates with B, A alters B's state of mind, and B's reaction

---

[12] The game unfolds as follows: nature chooses some point in the color space, according to some fixed probability distribution. S knows nature's choice, but r does not. Then s is allowed to send one of its messages to r. R in turn picks a point in the color space. S and r maximize utility if they maximize the similarity between nature's choice and r's interpretation. The only assumption is that the relevant similarity is a monotonically decreasing function of the distance in the color space between nature's choice and r's response.
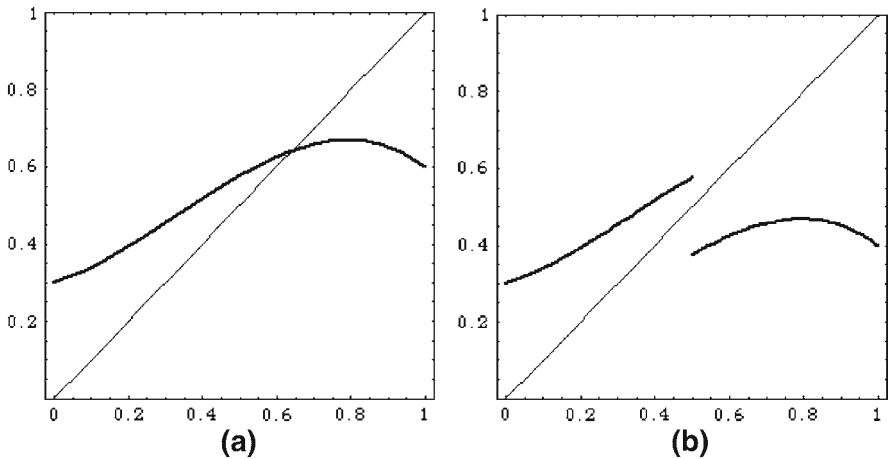
Fig. 4 **a** The fixpoint illustrated for a one-dimensional space. **b** Fixpoints may not exist if the function is not continuous

will, in turn, change A's state of mind.[13] In more formal terms, communication can be described with the help of *semantic reaction functions* from (x_1, x_2) to (y_1, y_2) in the product space $C = C_1 \times C_2$. In our framework, the semantic reaction function represents the inherent bi-directional coupling of production and comprehension processes: a point common to much recent research on pragmatics and radical semantics (Dekker and van Rooij 2000; Pickering and Garrod 2004; van Benthem 2008; Parikh 2010).

As a first abstraction, we assume that the reaction functions are continuous: that is, small changes in the communication will result in small changes in the reaction. We will show how such continuous functions can be approximated by discrete structures.

Now that all the ingredients are in place, we can simply remind readers of one of the most fundamental results of analysis: Brouwer's (1910) theorem, whereby each continuous mapping of a convex, compact set onto itself has at least one fixpoint. The continuous map we are concerned with in the present context is the semantic reaction function mapping the product space C onto itself. Note that C can be the product of several individual spaces and not just two. Figure 4a illustrates the fixpoint theorem for a function mapping a one-dimensional space onto itself, while Fig. 4b shows the necessity of the continuity assumption.

What Brouwer's theorem tells us is that, no matter the content of individual mental representations, so long as such representations are well shaped and the communicative device plastic enough to preserve the spatial structure of concepts, there will always be at least one point representing a meeting of minds.

So far we have not explicitly considered the role of language in how fixpoints are achieved. Since people are not telepathic, the mapping between individual conceptual spaces must somehow be mediated. Language is the primary mediator, although gestures and other such tools can also be used. Using language, the speaker maps his

---

[13] Note that at this stage, we are not modelling communication as such, only its effects on mental spaces.
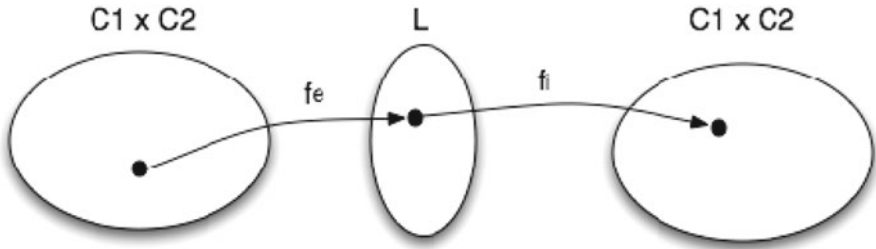
Fig. 5 A semantic reaction function maps points in $C_1 \times C_2$ to $C_1 \times C_2$ via L, where f_e is the expression function and f_i the interpretation function

mental space onto some linguistic entities from a language L, and the hearer in turn maps these expressions onto his mental space. Linguistic communication between two individuals with a product mental space $C_1 \times C_2$ is the composition of a function from C_1 to L and another function from L to C_2. This composition results in a modification of C_2: that is, a change of the hearer's mind. Put simply, linguistic communication results in a change from (x_1, x_2) to (x_1, y_2) in the product space.[14] In similar fashion, when the hearer responds, the mapping from C_2 via L to C_1 results in a move from (x_1, y_2) to (y_1, y_2) in the product space (Fig. 5).

Others' communicational intentions can sometimes be rationally anticipated, leading to a kind of instantaneous adjustment of the interlocutors to a fixpoint, as in a standard communication game. In general though, communication tends to be more myopic or at least adaptive, leading to a sequence of partial adjustments that appear to be the rule rather than the exception in dialogue-based communication (Clark 1992; Pickering and Garrod 2004). As an example of an elementary adaptive process, consider once more the case of pointing, and let C_1 and C_2 be the visual spaces of two individuals A and B. For simplicity, assume that $C_1 = C_2$. In the simplest case, the semantic reaction function starts from (x, y), where x is the point to which A is pointing, and y is the current position of A, which B is attending to in order to see the pointing direction of A. The reaction function maps the initial (x^*, y^0) to (x^*, y'), $(x^*, y'')\ldots(x^*, y^* = x)$, meaning that B is following A's gaze to the point x. The resulting fixpoint is the one at which A and B attend to the same object.

Clark's example of definite reference provides a more complex illustration of the nature of semantic reaction functions and the achievement of fixpoints. Cases in which conversation participants can jump directly to a commonly agreed reference point are relatively rare. Reference is usually achieved through a sequence of steps. The speaker must choose one out many possible utterances—typically a noun phrase—to initiate the process. The initial move may be mapped by the addressee into a state of mind inadequate to selecting the reference, in which case a sequence of adjustments will be triggered, requiring moves such as expanded descriptions, elaborations on or repairs to the original phrase, attempts to verify that the addressee has understood the reference, etc., until signals of mutual agreement are produced.

---

[14] In fact x_1 is altered too, since the expectations of the speaker change, particularly if the speaker has a *theory of mind* and can predict the listener's reactions.

The semantic reaction function need not be the result of an agent's explicit representation of the mental space of her interlocutor. As Pickering and Garrod (2004) have argued convincingly, many dialogic moves are primed automatically by the conversational context, favoring the convergence of interlocutors with a remarkable cognitive efficiency. An example of such automatic reactions is the tendency of listeners to complete the utterances of the speaker before she has finished producing them. Balzac's M. Grandet shrewdly exploited this automatism, using simulated babbling in the midst of commercial conversations to extract involuntary information from his interlocutors.

Brouwer's theorem depends on the reaction function being continuous. What is the meaning of continuity in the communication context? It is important to remember that the mental spaces of the interlocutors are based on similarity, which provides the spaces with their metric structure: in mental spaces, "near to" means "similar to". Continuity has a well-known nearness-preserving property, which, in our framework, becomes similarity preserving. In short, language preserves similarity in mental spaces.[15]

The mapping from C_1 to C_2 need not go via a discrete language, either verbal or signed. Other kinds of communication can be used, such as gestures, mimicry, and other such visual tools. Using these means, people can construct continuous functions between their mental spaces. The specific shape of such continuous functions will depend on a variety of pragmatic factors.

After this informal mathematical detour, our central claim should be clear: whenever it is important to reach a meeting of minds, convex mental representations provide the background for language to deploy most of its power. We see this as an indirect explanation of why concepts are generally convex.[16] We are not claiming that convex representations are faithful representations of the world—only that, since they are effective, one should find them quite widespread. In fact, one should expect to find convex representations even in cases where they are biased: seeing a non-convex world with convex spectacles might be a peculiar bias arising from selective pressures towards effective communication.

Brouwer's theorem proves the existence of fixpoints for any continuous function mapping a compact and convex space onto itself. Such a function is cognitively demanding, both in terms of memory and communication requirements. Our claim that the semantic reaction function is continuous seems at odds with the discreteness of language.

---

[15] The notion of nearness, introduced by F. Riesz, provides a geometrically more intuitive foundation for continuity than the classical $\varepsilon - \delta$ definition (see Cameron et al. 1974). The definition is simple for metric spaces, and can be generalized to arbitrary topological spaces. A point $x \epsilon X$ is near a subset $A \subseteq X$ if every neighbourhood of x contains a point in A. It can be shown that continuous maps preserve the nearness relation: i.e., if x is near A, $f(x)$ is near $f(A)$. This avoids the usual counterintuitive difficulties associated with the traditional contravariant definition of continuity.

[16] We view this argument as an inference to the best explanation. We do *not* view convexity as a necessary condition for reaching a fixpoint. We are aware that, for many phenomena, one may use weaker assumptions than employed in this paper: e.g., generalized convexity, purely ordinal topologies, multi-valued functions. We only want to stress the strong association between the convexity of concepts and the existence of fixpoints. We thank a reviewer for pointing out this ambiguity.

Indeed, the continuity of the semantic reaction function may seem too demanding an assumption, in light of the inherent discreteness of the lexicon. The challenge of reconciling lexical discreteness and, more generally, the discrete symbolic nature of language with the continuous, spatial nature of conceptual systems has been fundamental in recent attempts to re-unify lexical semantics (e.g. Jackendoff 2002). To understand that there is no real conflict, consider a fundamental result of algebraic topology: the simplicial approximation theorem, showing that any continuous function between two Euclidean spaces can be approximated by a mapping between the vertices of some appropriate triangulation of the spaces (see Appendix 1). As we have already shown, Voronoi tessellations provide a simple model for how categorization subdivides a conceptual space into convex sets. We have also described how the Voronoi diagram has a dual graph, which is a set of triangles generated by joining contiguous prototypes: that is, the Delaunay triangulation. This suggests that prototypes generate a basic triangulation of conceptual spaces, in which they play the role of simplicial vertices. A natural interpretation is that prototypes can provide the corners of the simplicial approximation of a continuous map between mental spaces. The correspondence between prototypes and words (or other lexical elements) then explains how language can serve as a mediator between conceptual spaces, approximating a continuous function. The result is a great economy in the cognitive resources needed to memorize and process such a function. For language users, the approximation boils down to remembering and communicating about the prototypes of a Voronoi decomposition of the space. In this way, the approximation can serve as a bridge between the discreteness of language and the continuity of the semantic reaction function. The bottom line is that the mechanisms of language and linguistic categorization are sufficient to approximate continuity with economic discretization. The convexity of spaces plays two important roles here: it ensures triangulability, and it allows reconstructing the behavior of the approximated function as convex combinations of the values of the approximated function in the correspondences of vertices.

It may be necessary to change the grain of the triangulation of the spaces in order to grant a simplicial approximation. Interestingly, human categorization systems have different levels of granularity, corresponding to different prototypes (Rosch 1975, 1978). As noted in Appendix 1, such an approximation between two triangulated spaces may only be achievable by further triangulating such spaces. Moving between levels of categorization may ensure that finer triangulations can be constructed and a simplicial approximation achieved. Note that, just as categories can be refined locally, not all simplexes of a simplicial complex need to be further triangulated to achieve the required degree of decomposition. Given a complex K, one can leave unchanged a subcomplex P while further triangulating another subcomplex—generating what is called a subdivision relative to P.

3.3 Shaping the semantic reaction function

Communication is always for some reason. We have postponed until now the issue of how motivation, and the stakes of communication, shape the semantic reaction func-

tion. Given what we have presented, it might seem that the semantics we propose is purely mentalistic. Reality enters via the payoffs of communication. If meaning is not properly aligned with reality, then communication may incur costs. Reality is why not all of our wishes come true—or, as Philip K. Dick expresses it: "reality is that which, when you stop believing in it, doesn't go away."

Reality enters communication when we use indexicals. A paradigmatic case is pointing in combination with saying "this" or that." As we discussed in the introduction, pointing is a way of coordinating our visual spaces, via reference to an external world.

Schelling's (1960) coordination game has often been used to introduce simple forms of semantic equilibria (see also Lewis 1969). In such games, all equilibria have the same payoff; a selection must be made by resorting to external factors, such as conventions or perceptual salience. Of course, reality may enter such games by making the payoffs of some equilibria superior. To make a Schelling-style example: imagine that you have agreed to a meeting in a very large square. There are infinitely many equilibria corresponding to the different locations in the square. However, it happens to be raining, and only one spot in the square is protected from rain. The obvious equilibrium to be selected is the one with the payoff of being protected from the rain. In this way, reality payoffs can be used to select equilibria in communication games. A related example once more involves definite reference. While doing some electric repair work, Mary asks Robert to bring her the screwdriver. There are two identical screwdrivers in the room: one close to Mary and Robert, the other much more distant. Obvious reasons of payoff dominance suggest the closest hammer as the referent.[17]

In other cases, reality may disrupt meetings of mind. Small children sometimes confuse height attributes with age. A mother's description of an adult A as older than an adolescent B may generate, in the mind of her child, the conviction that A is taller than B. The child's signs of assent may convince the mother of mutual understanding—but as the tall adolescent B shows up, mutual understanding may collapse.

In Jäger and van Rooij (2007) example, explicit payoffs were introduced for the success of the communication; the payoffs generated the best response function that determined the fixpoint. Considering the pragmatic factors that determine communication payoff leads naturally into communication games where speech acts become moves that modify the conversational playground.

Another advantage of our approach is that it allows one to consider a wide variety of types of communicative interaction, corresponding to different game types. For example, one might want to distinguish between coordination and negotiation games. In coordination games, such as the Jäger and van Rooij color game, the participants share common interests. In negotiation games, the participants have an interest in reaching agreement, but they have diverging interests on which agreement to reach.

Consider a semantic negotiation game, in which the interlocutors have an interest in agreeing on a meaning, but each of them wants agreement on different things. This resembles a contractual negotiation. The agents may have a partial overlap in representing a certain requisite concept, but in order to reach agreement they need to

---

[17] Parikh (2010) shows more complex examples of how payoff dominance may select meaning.

negotiate a sufficiently common meaning.[18] The process is obviously fallible. One can find many examples of contractual breaches that originate from different meanings associated with agreed contractual terms. In a famous court case, the contractors could not agree on the meaning of the "chicken" that was to be delivered (Frigaliment Imp. vs. B.N.S. Int'l Sales, 190 F. Supp 116—S.D.N.Y. 1960).

If the mental spaces of the interlocutors diverge widely, more radical communication methods must be applied in order to achieve a meeting of minds. This is where metaphors become powerful tools. By using a metaphor that exploits a shared domain, the speaker can convey information about another domain that has no or only vague correspondence in the mental spaces of the listeners. (We provide a treatment of metaphor in Sect. 4.3.) For example, if one wants to express an emotional experience that goes beyond the experiences of one's listeners, a metaphorical description is often the only available resource.

Some final notes on the fixpoint argument: We have noted the relationship between our approach and the literature on communication games. There are, however, potentially important differences. While fixpoints are central to defining game equilibria, fixpoints and game equilibria are not necessarily coincident. Some fixpoints may not correspond to game equilibria—at least to equilibria as standardly referred to in the communication games literature. The semantic reaction function is not necessarily a best response function: as Pickering and Garrod (2004) show convincingly, many conversational reactions are, to large extent, automatic: primed without any evaluation of alternatives. Indeed many adaptive processes may be conducive to fixpoints that are not best-response equilibria. While we will often make use of examples from communication games, it should be understood that fixpoint semantics, as we use it, is a broader notion than fixpoints in game semantics.

We should avoid implying that fixpoints necessarily mean successful communication, at least if "successful" corresponds to "optimal". Indeed, communication can be effective (i.e., generate a meeting of minds) *without* generating optimal payoffs in the activities it is intended to support.

Finally, a meeting of minds need not mean an identity of representations between agents. It might even be impossible to conceive a perfect alignment of representations among conversation participants (Parikh 2010)—even though some alignment is part of the convergence process in most conversations. (See again Pickering and Garrod 2004.) A strength of our approach is that it can encompass communicative agreement between participants even when their conceptual spaces differ. Homeomorphic spatial properties—rather than content of conceptual spaces—may, in many cases, be enough to sustain communication. What makes communication possible is the capacity to establish similarity-preserving mappings between the conceptual spaces of the participants.

---

[18] The convexity of their conceptual spaces may also help in reaching a solution.

4 Compositionality

Up to this point, we have treated the emergence of meaning as the holistic process of finding a fixpoint with a single utterance. All real communication is an articulated process that involves the composition of simple meaning components into larger wholes. Students of language have addressed this issue by different strategies. The Fregean tradition has emphasized the combinatorial nature of meaning composition, whereby the meaning of a composite expression is fully determined by the meaning of its components and the rules for composing its structure. Students of metaphor have introduced non-Fregean meaning composition as mappings across semantic domains (Holyoak and Thagard 1996; Fauconnier and Turner 1998; Gärdenfors 2000). Finally, conversation analysis has emphasized a sequential composition of meaning through the accumulation of a *common ground* (Clark and Schaefer 1989). In this section, we suggest ways to reconcile and unify these traditions.

4.1 Direct composition

Compositionality is a fundamental semantic property. Compositionality emerges directly from our framework of space and functions. To give a trivial example: the meaning of "blue rectangle" is defined as the Cartesian product of the "blue" region of color space and the "rectangle" region of shape space (which we leave undefined here[19]). Note that the product of any compact and convex sets will again be a compact and convex set: the structural properties of conceptual spaces are preserved under this basic composition operator.
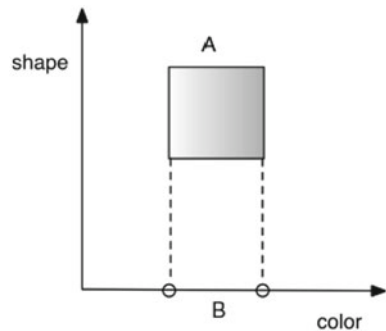
Not only are topological properties preserved but also the continuity of functions. If functions $f : A \to X$ and $g : B \to Y$ are both continuous, then the product function $h = (f, g) : A \times B \to X \times Y$ is also continuous. The composition of continuous functions $(g \circ f)$ is likewise continuous. This allows one to concatenate functions while preserving their basic properties.[20] Taken all together, the consequence is that compositionality can preserve, at a higher level of aggregation, the fixpoint properties of simpler meaning components. Compositionality can also help fine tune the grain of the fixpoint approximation by allowing finer decompositions of a conceptual space. A policeman helping a tourist find a certain building can progressively refine his description of the surrounding buildings by compositionally adding attributes—the yellow building with a large iron gate on the right of the post office—making smaller and smaller the set of buildings under consideration. In this way, compositionality enriches the set of communication moves available to achieving a meeting of minds.

The blue rectangle conceptual region can be decomposed into its generating regions "blue" and "rectangle" via projection (in turn is a continuous function) from the product space to its component spaces. Once again, the compactness and convexity of the blue rectangle region are preserved under projection (Fig. 6).

---

[19] For an analysis, see Gärdenfors (2000, Sect. 3.10.1).

[20] In a totally different spirit, Lewis (1970) uses the compositionality of functions to analyse various linguistic categories.

Fig. 6 Projection on the color dimension

One can recursively create ever richer composite concepts that preserve the basic topological properties of the original conceptual spaces: there seems to be no upper bound. There *is* a lower bound to decomposition via projection. Either the projection is one-dimensional, so that no further projection makes sense, or it is an integral set of dimensions, which we will call a domain.[21] For example, an object cannot be given a hue without simultaneously giving it a brightness. A domain is not decomposable because it cannot be reconstructed as the product of lower dimensional projections.

The proposed analysis of meaning compositionality is not the same as the classical Fregean one. Compositionality is traditionally generated from the meaning of words or expressions, while on our account it is generated from domains and functions. Since each domain is associated with a class of words—e.g., the class of color words—composing domains generates a composite conceptual space, and the meanings of composite expressions can be located as regions of the composite space.

Our implicit assumption is that the domains of the product constructions are independent. The reality of linguistic usage shows that that the spaces associated with composite expression are not fully independent; some pre-processing must take place before they can be properly composed. As an example of pre-processing, we will consider some cases of modifier-head composition.

4.2 Modifier—head composition

In the simplest cases—such as "blue rectangle", where "blue" is the modifier and "rectangle" is the *head*—the two associated domains (color and shape) can safely be assumed to be independent. However, this is rare in actual language use. More commonly, our knowledge of the space associated with the head may affect our representation of the modifier. White wine is not white; a large squirrel is not a large animal; and a thick forest does not compare to thick hair. In all these cases, some preprocessing

---

[21] The notion of integral dimensions and their connection to domains is presented in Gärdenfors (2000, Sect. 1.8).
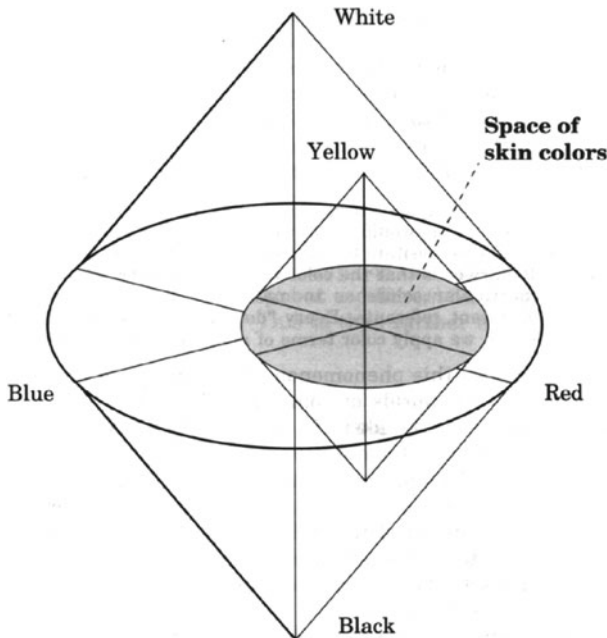
Fig. 7 The subspace of *skin colors* embedded in the full *color spindle* (from Gärdenfors 2000, p. 121)

of the representation of the modifier space seems to be required to adapt it to one's knowledge of the head space.[22]

Some properties cannot be defined independently of other properties at all. Consider "tall", which is connected to the height dimension but cannot be identified with a specific region in this dimension. To see the difficulty, consider that a chihuahua is a dog, but a tall chihuahua is not a tall dog. "Tall" presumes some contrast class given by some other property: things are not tall in themselves but only in relation to some other class of things. For a given contrast class Y—say the class of dogs—the region $H(Y)$ of possible heights of objects in Y can be determined. A particular Y can be said to be a tall Y if it belongs to the upper part of the region $H(Y)$.

For a contrast class such as skin color, one can map out the possible colors on the color spindle. This mapping will determine a subset of the full color space. If the subset is given the same geometry as the full color space, one obtains something like Fig. 7. In the smaller spindle, the color words are used in the same way as in the full space, even if the hues of the color in the smaller space don't match the hues of that color in the complete space. Thus, "white" is used to describe the lightest shades of skin, even though "white" skin is actually pinkish; "black" refers to the darkest shades of skin, even though most "black" skin is brown; and so on.

Provided that the head and modifier spaces are compact and convex regions of metric spaces, a way always exists of re-scaling the distances in the modifier space to fit

---

[22] Note that it is not the head but the modifier that is modified before the composition of meanings takes place.

Fig. 8 A radial projection
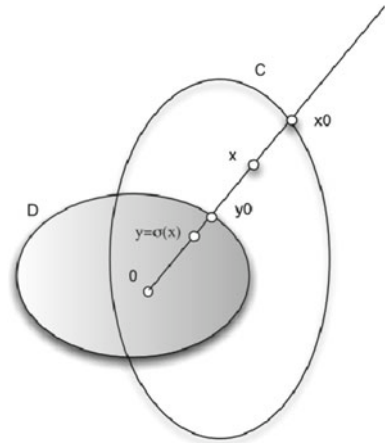
the constraints of the head space in a one-to-one correspondence. In this way, all color words are available to characterize skin color. *Radial projection* (Berge 1997) provides a natural conceptual tool to model such contextual re-scaling effects.[23] Consider two convex sets C and D defined within a space X and sharing an interior point 0 (taken arbitrarily as the *origo*). A radial projection is a mapping that establishes a one-to-one correspondence between the points in C and those in D, as well as between the frontier points of C and those of D. Such a correspondence re-defines the distance between the points in C and the *origo* in terms of the distance between the corresponding points in D. Figure 8 shows an example of correspondence between points on the boundaries (x_0 and y_0), and points in the interior (x and y).

It can be shown that a radial projection establishes a homeomorphism between two convex sets. So long as two sets are convex and compact and have a common interior point, such homeomorphism always exists (Berge 1997, p. 167).

This allows us to formulate a general principle: if the region of space representing the head contains a point shared with the space representing the modifier, this point can be taken as the *origo* of a transformation of the modifier space. The radial projection function tells one how to import structure from other domains. So long as concepts are convex and compact, such a function always exists. Radial projection is a continuous function—so, again, all transformations preserve nearness!

If the head and modifier share only some dimensions, the modifier ("pet" in "pet fish") or the head ("stone" in "stone lion") is projected onto the shared subspace and then expanded into a shared space by inverse projection. With "stone lion", the representation of stone includes the property "non-living," while "living" is presumed by many of the domains of "lion". These domains—e.g., sound, habitat, behavior—cannot be assigned any region at all. By and large, the only domain of "lion" that is compatible with stone is the shape domain. Consequently, the meaning of "stone lion" is an object made of stone in the shape of a lion.

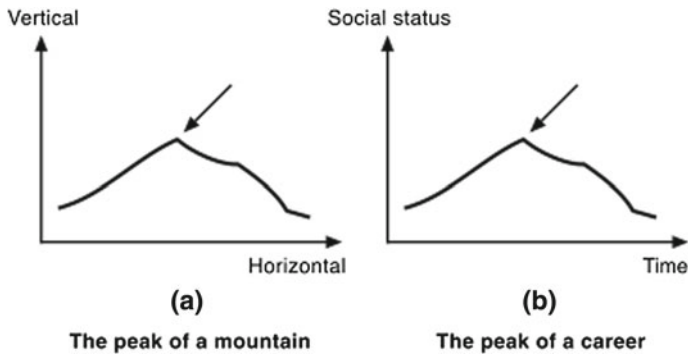[23] See Appendix 2 for a more rigorous definition of radial projection.

Fig. 9 **a** The literal meaning of "peak" **b** Metaphorical re-interpretation of (from Gärdenfors 2000, p. 177)

Gärdenfors (2000, p. 122) formulates the following rule for the meaning of the composition of a head *D* with modifier *C*: "The combination *CD* of two concepts *C* and *D* is determined by letting the regions for the domains of *C*, confined to the contrast class defined by *D*, replace the corresponding regions for *D*." One can see how this principle follows as a consequence of the account presented in this section. The head of a composed expression introduces a context that influences the meaning of the modifier. We will discuss more general context effects in Sect. 4.4.

4.3 Metaphorical composition

Even if the head and the modifier do not share *any* dimensions, one can still create a mapping between the domains by exploiting their convexity and compactness. Indeed, an important consequence of radial projection is that any two convex, compact spaces can be mapped via homeomorphism. This permits the creation of metaphors as the transfer of structure from one domain to another. Once the transfer is made, one is back to the previous modifier action.

Consider the expression "the peak of a career". The literal meaning of "peak" refers to a structure in physical space: namely, the vertically highest point in a (large) horizontally extended object, typically a mountain. This presumes two spatial dimensions, horizontal and vertical (see Fig. 9a).

A career is an abstract entity with no spatial location. How can a career have a peak? What happens is that the same geometrical structure from the mountain is applied to a two-dimensional space: the horizontal spatial dimension is mapped onto the time dimension (what point the person is in her career), and the vertical spatial dimension is mapped onto a dimension of social status (see Fig. 9b). The latter dimension is typically conceived of as vertical: one talks about somebody having a "higher" rank, "climbing" in the hierarchy, etc. (see Lakoff and Johnson 1980).

A metaphor does not come alone: it not only compares two concepts but also the structure of two complete domains. Once two domains have been connected via metaphor, this connection may serve as a generator for new metaphors based on the same kind of relations (see also Lakoff and Johnson 1980; Tourangeau and Sternberg 1982;

Gärdenfors 2000, Sect. 5.4). In brief, metaphorical mappings involve whole systems of concepts.

We have presented three ways of composing a modifier and a head. The first is plain compositional product construction. The second presupposes an adaptation of the modifier space, as a radial projection (via a gauge function) onto the head space. The third, involved in metaphor, requires an additional homeomorphic mapping between spaces. All three cases presume that mental spaces are partitioned into domains. The topological properties of the spaces are preserved, preparing them for further composition—creating finer approximations of a continuous semantic reaction function.

The composition discussed in Sect. 4.1 is really a special case of the composition in Sect. 4.2, which in turn is a special case of the composition in Sect. 4.3. The first sort need not modify existing spaces. The second sort modifies spaces that are naturally overlapping. Finally, the metaphors in Sect. 4.3 require establishing homeomorphic correspondences between the different (non-overlapping) spaces. In consequence, we expect that each of the three levels of composition will in turn require increasing cognitive processing.

4.4 Context and common ground

The composition of meaning usually takes place in multiple steps in larger communication structures such as conversations. A full account of the composition of meaning should therefore consider the sequential process through which meaning is determined.

Conversation analysis has singled out two main processes of temporal composition. One is the aforementioned accumulation of common ground, including the conversational use of context (Lewis 1979; Stalnaker 1979; Clark 1992). The other is the interactive alignment of representations (Pickering and Garrod 2004). These two processes can be clearly distinguished in our framework: the first is a progressive restriction within a given representation, creating finer partitions; the other modifies the salience of the conceptual dimensions carried by the topic of the conversation.

A typical move in a communication game that proposes to enlarge the common ground is an assertion (Clark 1992; Warglien 2001). Stalnaker (1979, p. 323) writes: "…the essential effect of an assertion is to change the presuppositions of the participants in the conversation by adding the content of what is asserted to what is presupposed. This effect is avoided only if the assertion is rejected." If the listener's countermove is accepting the assertion, then both parties contract their mental spaces (and expand their belief states). If, on the other hand, the listener rejects the assertion, another move will need to be attempted (for examples, see Clark 1992).

Sentences can affect the common ground through a reduction of the possibilities left open (Stalnaker 1979; Heim 1983; Clark 1992). Stalnaker assumes that the goal of a conversation is to converge towards a smaller set of possible worlds. We do not need the assumption of a strong shared representation on our approach. So long as the mental spaces of the participants satisfy the necessary topological properties, then the structure of the spaces, together with contraction, guarantees convergence. More precisely, if the product space of the mental space C is compact, and the reaction function f : $C \rightarrow C$ is continuous, then:

If, for any n, $f(C) \supseteq f^2(C) \supseteq \cdots \supseteq f^n(C)$ (repeated applications of f to its range generate a decreasing sequence); and, for each x, $f(x) \neq \emptyset$, then there exists a compact subset K of the product space such that $f(K) = K$.[24]

If the communication game constantly contracts the mental spaces of the participants, convergence is (given compactness) intuitively unavoidable. The typical rules of conversation, as set down by Grice (1975), are communicational institutions that help maintain the contractive nature of communication and thus ensure convergence. In brief, compactness, together with Grice's maxims, results in convergence towards the common ground. The condition "repeated applications of f to its range generate a decreasing sequence" corresponds to the maxim of quantity (always be informative), and the condition "$f(x) \neq \emptyset$" corresponds to the maxim of quality (never speak falsely).

Alignment (Pickering and Garrod 2004) is a different process, relying on the tendency, given the multi-dimensional representation of a situation, for conversants to align, mostly by automatically primed processes, either whole representations or specific dimensions of representations that were previously introduced in the conversation. In more complex cases, domains may prime structurally related ones: i.e., a temporal reference frame can be primed by a spatial reference frame (Boroditsky 2000). In our framework based on dimensions and domains in conceptual spaces, the process of alignment can be seen naturally both as a process of selecting common domains and as one of determining the relative weight of specific dimensions. Our framework further allows modeling how alignment contributes to mutual agreement. A simple visual example will illustrate this last point. Two agents with different sets of prototypes (dots and stars in Fig. 10a) may generate different decompositions of a two-dimensional space (solid and dotted lines), such that their categorization of entities in that space will disagree most of the time. Nevertheless, as conversation proceeds, they may alter their representations: e.g., by priming a greater salience for the X axis. Changes in salience will result in changes of axis weights, altering the distance between points (Gärdenfors 2000). By adjusting the weight of the two dimensions, the agents will be able to align their categorization of points in the space such that they will come to agree most of the time (Fig. 10b).

A strong claim, taken from pragmatics, is that meaning is largely underdetermined until context is brought in. Our account suggests that context can affect the determination of meaning in two ways. One is by being added to the common ground as described above. The second is through the modification of representations. We have already discussed three ways in which such modification may happen: (1) the contrast effect whereby vague terms like "tall" become more determinate; (2) the metaphorical mappings that allow introducing additional content, especially when the speaker has reason to believe that the addressee lacks compatible representations; (3) the priming effects that make some dimensions more salient.

---

[24] The proof immediately follows from Theorem 8 in Berge (1997, p. 113).
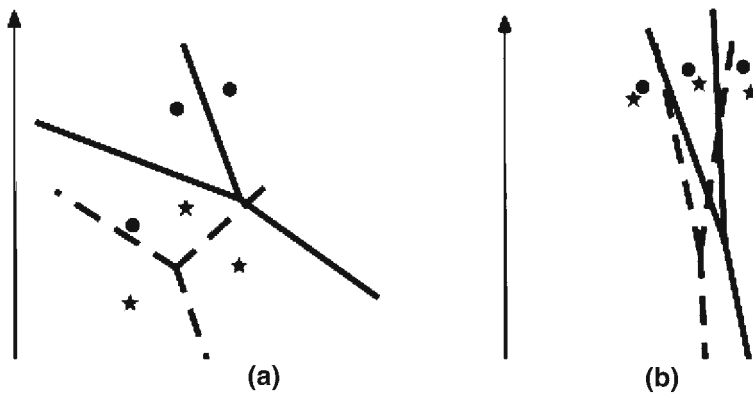
Fig. 10 Agreement by salience manipulation

5 Conclusion

This article presents a novel semantic theory, based on the meeting of minds. It puts more structure into communication games by exploiting the topological and geometric properties of individual mental spaces, modeled here in terms of conceptual spaces. Our approach emphasizes the shape of representations rather than their contents. One advantage, in contrast to cognitive semantics, is that we need not assume that the inter-locutors share identical mental spaces. This makes it possible to explain how people can misunderstand each other but still avoid losing contact altogether.

Putnam (1975) has criticized cognitive approaches to semantics, concluding that "meanings ain't in the head" and that external reality must be a component in any reasonable semantic theory. *Contra* Putnam's argument, our socio-cognitive framework shows that meanings *are* in the heads of language users, when their minds meet.[25] Since our minds can meet over total fantasies, reality is not required for meaning to arise or for communication to succeed. Reality enters the picture when there are stakes to the communication, such that success depends on the outcome of some communication game.

An advantage of our approach is that we are able to establish a connection between the discreteness of language resources and the continuity of thought. We have argued that language, in a broad sense including context and other pragmatic aspects, contains mechanisms to preserve the similarity of meanings. We show that language, with all its resources, is an efficient way of approximating the continuity of semantic mappings. We have focused on compositionality, but many other aspects of how linguistic structures preserve similarity should be investigated. We have shown how semantics can exist and a meeting of minds arise without language, in the pointing example. This is a great advantage to our approach: it unifies the different forms of communication and does not treat language as an exclusive carrier of meaning.
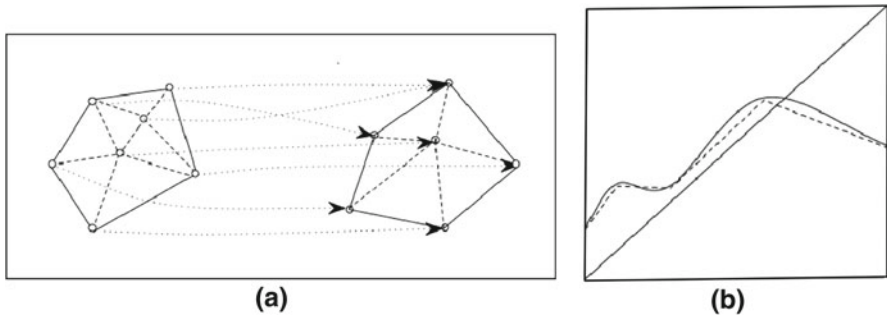
---

[25] See also Gärdenfors (2000, Sect. 5.7.)

Fig. 11 a Mapping triangulation. **b** A piecewise linear approximation of a one-dimensional function

It should be emphasized that our presentation here is just a framework and not a model. That said, our framework lends itself naturally to modeling. We have only considered linguistic structures in very general terms: some forms of composition applied mostly to noun phrases. The next step would be to connect different linguistic categories to particular mathematical operations.

Appendix 1: Simplicial approximation of a continuous function

This appendix briefly explains the concept of the simplicial approximation of a continuous function, which we used to relate fixpoints with the discreteness of language in Sect. 3. The basic idea can be stated quite simply, but it requires some preliminary definitions.[26] Let X be a convex compact set in a Euclidean space C. A triangulation K of X is a decomposition of that set in a finite set of simplexes, where a simplex is the set of convex combinations of n independent points in some m-dimensional Euclidean space. (Basically, it is an m-dimensional generalization of a triangle.) The triangulation further requires that two of the simplexes meet in (at maximum) one face or edge. The combinatorial structure thus generated constitutes a geometric simplicial complex. Although a geometric simplicial complex is not of itself a topological space, its points, topologized as a subspace of C, are a topological space: the polyhedron $|K|$. A simplicial map $f : |K| \rightarrow |L|$ between two polyhedra $|K|$ and $|L|$ is a function that maps the vertices of $|K|$ to the vertices of $|L|$, while preserving the simplexes. In other words, if $a^0, a^1 \ldots a^n$ are the vertices a simplex in K, then $f(a^0), f(a^1) \ldots f(a^n)$ are the vertices of a simplex in L.[27] Figure 11a provides a simple illustration. Note that

---

[26] In what follows, we heavily rely on the presentation in Maunder (1980).

[27] Notice that $f(a^0), f(a^1), \ldots f(a^n)$ need not all be different points: repetition is allowed.

it is required, if x is a convex combination of $a^0, a^1 \ldots a^n$, for $f(x)$ to be a convex combination of $f(a^0), f(a^1), \ldots f(a^n)$. It can be shown that f must be continuous.

A simplicial map is, as noted, a simplex preserving map. Simplicial maps have another very important property. Given two convex, compact sets X and Y and a continuous function $g: X \to Y$, there will always exist a simplicial map f approximating g, provided that X and Y can be triangulated at a sufficiently fine grain. By "f approximating g", it is meant that f is homotopic to g: put simply, that g can be obtained from a continuous deformation of f, a result known as the simplicial approximation theorem. The theorem proves that *any* continuous map can be approximated by a piecewise linear map. Figure 11b provides a simple example in which both X and Y are one-dimensional sets. The simplicial map f preserves the fixpoint properties of g and can be used to approximate the fixpoints of g.

An important proviso concerning is that, as mentioned above, it will not always be possible to make a simplicial approximation between two given triangulated spaces: specifically, where it is not possible to triangulate such spaces with a sufficiently fine grain. In other words, the triangulation of $|K|$ may be insufficiently fine grained to permit a simplicial approximation. That said, if |K| can be further subdivided, a simplicial approximation will always exist. The compactness of $|K|$ ensures such further triangulability. Only the subdivision of one part of a simplicial complex may be needed. Given a complex K, one can leave unchanged everything but a sub-complex P and further triangulate the sub-complex, generating a subdivision relative to P.[28]

Appendix 2: Radial projection

A gauge (also known as a Minkowski functional) of a convex set, with interior point 0 taken as the *origo*, is a generalized numeric function j defined as:

$j(x) = \inf\{t : t > 0, x \in tC\}$ if $x \in tC$ for at least one $t > 0$ (where tC is C inflated by factor t);

$j(x) = \infty$ otherwise.

It is easily shown (Berge 1997) that if $x \in C$, then $j(x) \leq 1$. (in particular, $j(x) = 1$ if x is on the boundary of the set.)

Consider convex sets C and D, both defined within a space X and possessing a common interior point 0 (taken arbitrarily as the *origo*). Let j and k be the gauge functions for C and D respectively. One can define the following function $\sigma: X \to X$:

$$y = \sigma(X) = \left(\frac{j(X)}{k(X)}\right) X$$

Such function (called a radial projection) establishes a correspondence between each point of C and a corresponding point in D:

$$k(y) = k\left[\frac{j(X)}{k(X)}\right] X = \frac{j(X)}{k(X)} k(X) = j(X)$$

---

[28] See Maunder (1980, Sect. 2.5.7).

References

Bates, E. (Ed.). (1976). *Language and context. The acquisition of pragmatics*. New York: Academic Press.

Berge, C. (1997). *Topological spaces*. Mineola: Dover.

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition, 75*(1), 1–28.

Brinck, I. (2004). The pragmatics of imperative and declarative pointing. *Cognitive Science Quarterly, 3*, 255–272.

Brinck, I., & Gärdenfors, P. (2003). Co-operation and communication in apes and humans. *Mind and Language, 18*, 484–501.

Brouwer, L. E. J. (1910). Über ein eindeutige, stetige Transformation von Flächen in sich. *Mathematische Annalen, 69*, 176–180.

Cameron, P., Hogkin, J. G., & Naimpally, S. A. (1974). Nearness: A better approach to continuity and limits. *The American Mathematical Monthly, 81*(7), 739–745.

Clark, H. (1992). *Arenas of language use*. Chicago: University of Chicago Press.

Clark, H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*, 259–294.

Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.

Dekker, P., & van Rooij, R. (2000). Bi-directional optimality theory: An application of game theory. *Journal of Semantics, 17*, 217–242.

Evans, V. (2006). Lexical concepts, cognitive models and meaning-construction. *Cognitive Linguistics, 17*, 491–534.

Fauconnier, G., & Turner, R. (1998). Conceptual integration networks. *Cognitive Science, 22*(2), 133–187.

Fitting, M. (2002). Fixpoint semantics for logic programming: A survey. *Theoretical Computer Science, 278*, 25–51.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science, 29*, 737–767.

Gärdenfors, P. (1997). Does semantics need reality? In *Does representation need reality?* (pp. 113–120). Austrian Society of Cognitive Science Technical Report 97-01, Vienna.

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

Gärdenfors, P. (2003). *How homo became sapiens: On the evolution of thinking*. Oxford: Oxford University Press.

Gärdenfors, P. (2007). Representing actions and functional properties in conceptual spaces. In T. Ziemke, J. Zlatev, & R. M. Frank (Eds.), *Body, language and mind: Embodiment* (Vol. 1, pp. 167–195). Berlin: Mouton de Gruyter.

Gärdenfors, P. & Osvath, M. (2010). The evolution of anticipatory cognition as a precursor to symbolic communication. In R. Larson et al. (Eds.), *Evolution of language: Biolinguistic approaches* (pp. 103–114). Cambridge: Cambridge University Press.

Gärdenfors, P., & Warglien, M. (to appear a). The development of semantic space for pointing and verbal communication. In J. Hudson, U. Magnusson & C. Paradis (Eds.), *Conceptual spaces and the construal of spatial meaning: Empirical evidence from human communication*. Oxford: Oxford University Press.

Gärdenfors, P., & Warglien, M. (to appear b). Using conceptual spaces to model actions and events, to appear in Journal of Semantics.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic coordination. *Cognition, 27*(2), 181–218.

Goldin-Meadow, S. (2007). Pointing sets the stage for learning language and creating language. *Child Development, 78*, 741–745.

Grice, P. (1975). Logic and conversation. In Peter Cole & Jerry L. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. 3, pp. 41–58). New York: Academic Press.

Harnad, S. (1990). The symbol grounding problem. *Physica D, 42*, 335–346.

Heim, I. (1983). On the projection problem for presuppositions. In Proceedings of the west coast conference on formal linguistics (Vol. II, pp. 114–125). Stanford, CA.

Holyoak, K. J., & Thagard, P. (1996). *Mental leaps*. Cambridge, MA: MIT Press.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States, 79*, 2554–2558.

Holmqvist, K. (1993). *Implementing cognitive semantics*. Lund University Cognitive Studies 17, Lund.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Jäger, G. (2007). The evolution of convex categories. *Linguistics and Philosophy, 30*, 551–564.

Jäger, G. (2010). Natural color categories are convex sets. In *Logic, language and meaning, LNCS 6042* (pp. 11–20). Berlin: Springer.

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

Johannesson, M. (2002). *Geometric models of similarity*. Lund University Cognitive Studies 87, Lund.

Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy, 72*, 690–716.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago, IL: The University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: The University of Chicago Press.

Langacker, R. W. (1986). An introduction to cognitive grammar. *Cognitive Science, 10*, 1–40.

Langacker, R. W. (1987). *Foundations of cognitive grammar* (Vol. 1). Stanford, CA: Stanford University Press.

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

Lewis, D. (1970). General semantics. *Synthese, 22*, 18–67.

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic, 8*, 339–359.

Maunder, C. R. F. (1980). *Algebraic topology*. Cambridge: Cambridge University Press.

Mervis, C., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*, 89–115.

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 54–65.

Okabe, A., Boots, B., & Sugihara, K. (1992). *Spatial tessellations: Concepts and applications of Voronoi diagrams*. New York: John Wiley & Sons.

Parikh, P. (2000). Communication, meaning and interpretation. *Linguistics and Philosophy, 23*, 185–212.

Parikh, P. (2010). *Language and equilibrium*. Cambridge, MA: MIT Press.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169–190.

Putnam, H. (1975). The meaning of 'meaning'. In K. Gunderson (Ed.), *Language, mind and knowledge* (pp. 131–193). Minneapolis: University of Minnesota Press.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*, 192–233.

Rosch, E. (1978). Prototype classification and logical classification: The two systems. In E. Scholnik (Ed.), *New trends in cognitive representation: Challenges to Piagets theory* (pp. 73–86). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schelling, T. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Pres.

Selten, R., & Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. *Proceedings of the National Academy of Sciences, 104*(18), 7361–7366.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*, 1317–1323.

Skyrms, B. (1998). Salience and symmetry-breaking in the evolution of convention. *Law and Philosophy, 17*, 411–418.

Stalnaker, R. (1979). Assertion. *Syntax and Semantics, 9*, 315–332.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tourangeau, R., & Sternberg, R. J. (1982). Understanding and appreciating metaphors. *Cognition, 11*, 203–244.

van Benthem, J. (2008). Games that make sense: Logic, language and multi-agent interaction. In K. Apt & R. van Rooij (Eds.), *New perspectives on games and interaction* (pp. 197–209). Amsterdam: Amsterdam University Press.

Varzi, A., & Warglien, M. (manuscript). Indeterminate contracts and semantic indeterminacy.

Warglien, M. (2001). *Playing conversation games*. Paper presented at the 2001 Wittgenstein society symposium, Kirchberg.